



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FLORESTAIS

**JEFERSON PEREIRA MARTINS SILVA**

**PROGNOSE DA PRODUÇÃO FLORESTAL UTILIZANDO SISTEMA *NEURO-FUZZY* E *RANDOM FOREST***

JERÔNIMO MONTEIRO – ES  
FEVEREIRO – 2018

**JEFERSON PEREIRA MARTINS SILVA**

**PROGNOSE DA PRODUÇÃO FLORESTAL UTILIZANDO SISTEMA *NEURO-FUZZY* E *RANDOM FOREST***

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Ciência Florestal na Área de Concentração Ciência Florestal.

Orientadora: Mayra Luiza Marques da Silva  
Coorientador: Adriano Ribeiro de Mendonça

JERÔNIMO MONTEIRO – ES

FEVEREIRO – 2018

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Setorial Sul, Universidade Federal do Espírito Santo, ES, Brasil)  
Bibliotecário: Felício Gomes Corteletti – CRB-6 ES-000646/O

---

S586p Silva, Jeferson Pereira Martins, 1992-  
Prognose da produção florestal utilizando sistema neuro-fuzzy e  
random forest / Jeferson Pereira Martins Silva. – 2018.  
67 f. : il.

Orientador: Mayra Luiza Marques da Silva.

Coorientador: Adriano Ribeiro de Mendonça.

Dissertação (Mestrado em Ciências Florestais) – Universidade  
Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Produção florestal. 2. Inteligência artificial. 3. Mensuração  
florestal. I. Silva, Mayra Luiza Marques da. II. Mendonça, Adriano  
Ribeiro de. III. Universidade Federal do Espírito Santo. Centro de  
Ciências Agrárias e Engenharias. IV. Título.

CDU: 630

---

**PROGNOSE DA PRODUÇÃO FLORESTAL UTILIZANDO SISTEMA NEURO-FUZZY  
E RANDOM FOREST**

**Jeferson Pereira Martins Silva**

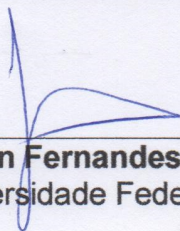
Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do Título de Mestre em Ciências Florestais na Área de Concentração Ciências Florestais.

Aprovada em 28 de fevereiro de 2018.



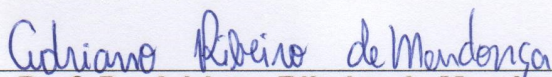
---

**Prof. Dr. Antonio Almeida de Barros Junior** (Examinador externo)  
Universidade Federal do Espírito Santo



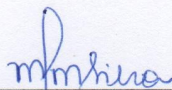
---

**Prof. Dr. Gilson Fernandes da Silva** (Examinador interno)  
Universidade Federal do Espírito Santo



---

**Prof. Dr. Adriano Ribeiro de Mendonça** (Coorientador)  
Universidade Federal do Espírito Santo



---

**Profª. Drª. Mayra Luiza Marques da Silva** (Orientadora)  
Universidade Federal de São João del-Rei

*“ A utopia está lá no horizonte. Me aproximo dois passos, ela se afasta dois passos. Caminho dez passos e o horizonte corre dez passos. Por mais que eu caminhe, jamais alcançarei. Para que serve a utopia? Serve para isso: para que eu não deixe de caminhar.”*

Eduardo Hughes Galeano

***Dedico***

*Aos meus pais Arlene e Manoel Vinícius*

*Aos meus irmãos Jéssica e Vinícius*

## **AGRADECIMENTOS**

A Deus, por me guiar e dá forças para sempre seguir com meus objetivos.

Aos pais, Arlene e Manoel Vinícius, por mostrar, a mim e aos meus irmãos, o significado de família.

Aos meus irmãos, Jéssica e Vinícius, pelo carinho incondicional e por sempre estarem ao meu lado.

Ao Emanuel (primo), pela boa convivência durante esses dois anos de mestrado.

À minha orientadora, Mayra Luíza Marques da Silva, por todo conhecimento passado e ajuda na elaboração desta dissertação.

Ao meu coorientador, Adriano Ribeiro de Mendonça, pelas contribuições nas correções para melhorias deste trabalho.

Ao professor Gilson Fernandes da Silva, pelos valiosos ensinamentos e o exemplo de profissional que é.

Ao professor Antônio Almeida de Barros Junior, pela prontidão em participar da banca como examinador.

Aos grandes amigos do laboratório LaMFlor, pelo bom convívio e o apoio durante a realização desta dissertação.

Aos amigos da república Batista House, Genilda, Kaíse, Mariana e Márcia, muito bom partilhar com vocês todos esses momentos.

Ao Centro de Ciências Agrárias e Engenharias, ao Programa de Pós-graduação em Ciências Florestais, à Universidade Federal do Espírito Santo e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.



SILVA, Jeferson Pereira Martins. **Prognose da produção florestal utilizando sistema *neuro-fuzzy* e *random forest***. 2018. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientadora: Mayra Luiza Marques da Silva. Coorientador: Adriano Ribeiro de Mendonça.

## RESUMO

O objetivo deste estudo foi avaliar o emprego das técnicas *Random Forest* (RF) e Sistema *Neuro-Fuzzy* (SNF) na prognose da produção florestal. Os dados utilizados foram provenientes de inventários florestais contínuos conduzidos em povoamentos de clones de eucalipto, localizados no sul da Bahia. O processamento dos dados foi realizado no software Matlab R2016a. Os dados foram divididos em 70% para de treinamento e 30% para validação. Os algoritmos usados para geração de regras no SNF foram *Subtractive Clustering* (SC) e *Fuzzy-C-Means* (FCM). O treinamento foi feito com o algoritmo híbrido (gradiente descendente e mínimos quadrados) com o número de épocas variando de 1 a 20. As funções de pertinências associadas às variáveis de entradas foram do tipo gaussianas e a função linear na de saída. Foram treinadas várias RF variando o número de árvores de 50 a 850 e o número de observações por folhas de 5 a 35. A modelagem da produção florestal de povoamentos clonais de eucalipto pode ser realizada com SNF e RF. Os algoritmos SC e FCM fornecem estimativas acuradas na projeção de área basal e volume. A RF apresentou estatísticas inferiores em relação a SNF para prognose da produção florestal. Ambas as técnicas são boas alternativas para seleção de variáveis empregadas na modelagem da produção florestal.

**Palavras-chave:** Inteligência artificial, *ensemble learning*, mensuração florestal.

SILVA, Jeferson Pereira Martins. **Prognosis of forest production using a neural-fuzzy system and random forest**. 2018. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientadora: Mayra Luiza Marques da Silva. Coorientador: Adriano Ribeiro de Mendonça.

### **ABSTRACT**

The aim of this study to apply the techniques Random Forest (RF) and Neuro-Fuzzy System (SNF) in the prognosis of forest production. The data used came from continuous forest inventories conducted in settlements of eucalypt clones, located in the south of Bahia. The data processing was performed in Matlab R2016a software. Data were divided into 70% for training and 30% for validation. The algorithms used to generate rules in SNF were Subtractive Clustering (SC) and Fuzzy-C-Means (FCM). The training was done with the hybrid algorithm (descending gradient and least squares) with the number of times varying from 1 to 20. The pertinence functions associated with the input variables were of the gaussian type and the linear output. Several RF were trained by varying the number of trees from 50 to 850 and the number of observations per leaves from 5 to 35. The modeling of the forest production of clonal eucalypt stands can be performed with SNF and RF. The SC and FCM algorithms provide accurate estimates of basal area and volume projection. The RF presented inferior statistics in relation to SNF for prognosis of forest production. Both techniques are good alternatives for the selection of variables used in the modeling of forest production.

**Keywordsn:** Artificial intelligence, ensemble learning, forest measurement.



## SUMÁRIO

<b>1 INTRODUÇÃO</b>	12
<b>2 OBJETIVOS</b>	14
2.1 Objetivo Geral	14
2.2 Objetivos Específicos	14
<b>3 REVISÃO BIBLIOGRÁFICA</b>	15
3.1 Modelagem do crescimento e da produção florestal	15
3.2 <i>Random Forest</i>	17
3.3 Redes Neurais Artificiais	21
3.4 Lógica <i>Fuzzy</i>	23
3.5 Sistema <i>Neuro-Fuzzy</i>	25
3.6 Técnicas de inteligência artificial aplicadas na modelagem do crescimento e produção florestal	29
<b>4 MATERIAL E MÉTODOS</b>	31
4.1 Descrição da área e do banco de dados	32
4.2 Seleção das variáveis que mais contribuem para prognose da área basal e do volume	33
4.3 Configurações das técnicas de inteligência artificial	33
4.3.1 Sistema <i>Neuro-fuzzy</i>	33
4.3.2 <i>Random Forest</i>	36
4.4 Prognose da produção florestal	36
4.5 Critérios para avaliação do treinamento e validação das técnicas de inteligência artificial	36
<b>5 RESULTADOS E DISCUSSÃO</b>	38
5.1 Área basal	38
5.1.1 Seleção das variáveis de entrada para projetar a área basal	38
5.1.2 Configurações selecionadas das técnicas de inteligência artificial utilizadas para projetar a área basal	40
5.1.3 Desempenhos das configurações das técnicas de inteligência artificiais selecionadas para prognose da área basal	41
5.2 Volume	47
5.2.1 Seleção das variáveis de entrada para prognose da produção volumétrica	47
5.2.2 Configurações das técnicas selecionadas para projetar o volume	49

5.2.3 Desempenhos das configurações das técnicas de inteligência artificiais selecionadas para prognose do volume .....	50
5.3 Prognose da área basal .....	55
5.4 Prognose do volume .....	57
<b>REFERÊNCIAS</b> .....	62

## LISTA DE FIGURAS

Figura 1 – Mecanismo de funcionamento da <i>Random Forest</i> .....	21
Figura 2 – Rede neural artificial.....	22
Figura 3 – Função gaussiana .....	24
Figura 4 – Função triangular .....	24
Figura 5 – Função trapezoidal.....	25
Figura 6 – Estrutura do Sistema <i>Neuro-Fuzzy</i> .....	26
Figura 7 – Fluxograma metodológico para realização da prognose da produção florestal com Sistema <i>Neuro-Fuzzy</i> e <i>Random Forest</i> . ....	31
Figura 8 – Variáveis mais influentes para prognose da área basal indicadas pela <i>Random Forest</i> (Figura 8A) e Sistema <i>Neuro-Fuzzy</i> (Figura 8B). ....	39
Figura 9 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) <i>versus</i> estimados no treinamento. ....	42
Figura 10 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) <i>versus</i> estimados na validação.....	43
Figura 11 – Dispersão dos erros em porcentagem em relação aos valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) no treinamento. ....	44
Figura 12 – Dispersão dos erros em porcentagem em relação aos valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) na validação. ....	44
Figura 13 – Frequência relativa dos erros em porcentagem para estimativas da área basal ( $\text{m}^2.\text{ha}^{-1}$ ) no treinamento. ....	46
Figura 14 – Frequência relativa dos erros em porcentagem para estimativas da área basal ( $\text{m}^2.\text{ha}^{-1}$ ) na validação.....	46
Figura 15 – Variáveis mais influentes para prognose do volume indicadas pela <i>Random Forest</i> e Sistema <i>Neuro-Fuzzy</i> . ....	48
Figura 16 – Valores observados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) <i>versus</i> estimados no treinamento. ....	51
Figura 17 – Valores observados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) <i>versus</i> estimados na validação. ....	52
Figura 18 – Dispersão dos erros em porcentagem em relação aos valores estimados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) no treinamento. ....	53
Figura 19 – Dispersão dos erros, em porcentagem, em relação aos valores observados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) na validação.....	53

Figura 20 – Frequência relativa dos erros, em porcentagem, para estimativas de volume ( $\text{m}^3.\text{ha}^{-1}$ ) no treinamento. ....	54
Figura 21 – Frequência relativa dos erros, em porcentagem, para estimativas de volume ( $\text{m}^3.\text{ha}^{-1}$ ) na validação. ....	55
Figura 22 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) <i>versus</i> estimados, dispersão dos erros em porcentagem em relação aos valores estimados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) e histograma da frequência relativa dos erros percentuais para os dados utilizados na prognose da área basal. ....	56
Figura 23 – Valores observados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) <i>versus</i> estimados, dispersão dos erros em porcentagem em relação aos valores estimados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) e histograma da frequência relativa dos erros percentuais para os dados utilizados na prognose do volume.....	57

## LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas para o conjunto de treinamento e validação do Sistema <i>Neuro-Fuzzy</i> e <i>Random Forest</i> .....	32
Tabela 2 – Estatísticas para avaliar o desempenho das técnicas de inteligência artificial.....	37
Tabela 3 – Configurações pré-selecionadas do Sistema <i>Neuro-Suzzy</i> e <i>Random Forest</i> .....	40
Tabela 4 – Estatísticas do treinamento e validação do Sistema <i>Neuro-Fuzzy</i> e <i>Random Forest</i> para projetar a área basal ( $\text{m}^2.\text{ha}^{-1}$ ).....	41
Tabela 5 – Configurações pré-selecionadas do Sistema <i>Neuro-Fuzzy</i> e <i>Random Forest</i> .....	49
Tabela 6 – Estatísticas do treinamento e validação do Sistema <i>Neuro-Fuzzy</i> e <i>Random Forest</i> para projetar o volume ( $\text{m}^3.\text{ha}^{-1}$ ).....	50

## 1 INTRODUÇÃO

As florestas plantadas são recursos renováveis que contribuem para o desenvolvimento do país e ao mesmo tempo, diminuem a pressão sobre as florestas nativas. No Brasil, são estimados 7,84 milhões de hectares de florestas plantadas, no qual, o pinus e o eucalipto representam mais de 92% da produção de madeira. Essa extensa área de florestas plantadas, aliada às espécies de rápido crescimento, foi responsável por gerar, no período de janeiro a agosto de 2017, uma receita de 5,5 bilhões de dólares na exportação de celulose, papel e painéis de madeira, que representa um aumento de 11,4% em relação ao mesmo período de 2016 (IBÁ, 2017).

Segundo Miranda, Junior e Gouvea (2015), as áreas de florestas plantadas no Brasil estão aumentando de forma rápida. Isto pode estar associado à crescente demanda nacional e internacional por produtos de base florestal. Sendo assim, o conhecimento da produção, em termos de volume por hectare, é imprescindível para garantir a continuidade do fornecimento de madeira de qualidade e aumentar a competitividade do setor.

Os modelos de crescimento e produção ajudam pesquisadores e gerentes de empresas florestais, a estimar rendimentos volumétricos futuros, explorar opções de gestão e alternativas silviculturais (VANCLAY, 1994). Estes rendimentos são obtidos comumente por meio de modelos estatísticos (BINOTI et al., 2015). Outra possibilidade de modelagem do crescimento e da produção de povoamentos florestais, que vem mostrando bons resultados, é a aplicação das técnicas de inteligência artificial, por exemplo, as Redes Neurais Artificiais (RNA).

As RNAs têm gerado maior acurácia com relação às técnicas estatísticas tradicionais, como observado nos estudos de Diamantopoulou (2005), para estimar volume de pinus, Binoti et al. (2013), para estimação da altura de povoamentos equiâneos de eucalipto, Binoti et al. (2014), para prognose em nível de povoamento de clones de *Eucalyptus*, dentre outros. Outras técnicas de inteligência artificial com potencial de aplicação na mensuração e manejo florestal são o Sistema *Neuro-Fuzzy* (SNF) e a *Random Forest* (RF).

O SNF é a combinação entre a RNA e a lógica *fuzzy*, em que a rede neural gera regras ou funções de pertinência para os sistemas *fuzzy* (SIMÕES; SHAW, 2007). O SNF tem demonstrado ser uma ferramenta eficiente para modelagem de

sistemas da vida real. Possuem boa interpretabilidade, são precisos e conseguem generalizar o conhecimento adquirido (SIMINSKI, 2017). Bons resultados foram encontrados em trabalhos da área florestal utilizando o SNF: Silva et al. (2007), para estimar o volume total de eucalipto, Silva et al. (2008), para estimar altura total de árvores de eucalipto e Vieira et al. (2018), para estimar o crescimento em diâmetro e em altura de árvores de eucalipto.

A RF é um método de aprendizagem de máquina que utiliza uma abordagem *bootstrap* para construir várias árvores de decisão. A RF pode ser usada para problemas de classificação e regressão (BREIMAN, 2001). Além disso, possuem um sistema de seleção de recursos na sua estrutura, utilizando somente as variáveis que realmente influenciam na resposta (CANO et al., 2017). São exemplos de aplicação da RF: classificação de imagens (CRACKNELL; READING, 2014; AKKOÇ; ARSLAN; KÖK, 2017; MASOUMI et al., 2017), seleção de atributos (JANITZA; TUTZ; BOULESTEIX, 2016; GOUDARZI, 2016; CHELGANI; MATIN; MAKAREMI, 2016) e aproximação de funções (CHERRY et al., 2015; TOOKE; NUNES; GÖRGENS, 2016; WANG et al., 2016; CANO et al., 2017).

Embora existam na literatura muitos estudos envolvendo o uso das RNA aplicadas às ciências florestais, trabalhos com a utilização de técnicas como o SNF e a RF são ainda incipientes, o que justifica a importância de estudos com aplicação dessas técnicas no setor florestal, principalmente no que se refere a estimação da produção florestal, visto que é uma das principais atividades do manejo.



## 2 OBJETIVOS

### 2.1 Objetivo Geral

Avaliar o emprego do Sistema *Neuro-Fuzzy* e *Random Forest* na prognose da área basal e do volume de povoamentos de eucalipto.

### 2.2 Objetivos Específicos

- Avaliar e selecionar dentre as variáveis que mais influenciam na estimacão da área basal e do volume.
- Configurar os parâmetros do Sistema *Neuro-Fuzzy* e *Random Forest*, para estimacão da área basal e do volume.
- Realizar a prognose da área basal com as técnicas Sistema *Neuro-Fuzzy* e *Random Forest*.
- Realizar a prognose do volume com as técnicas Sistema *Neuro-Fuzzy* e *Random Forest*, utilizando a área basal futura estimada.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Modelagem do crescimento e da produção florestal

Para um bom planejamento das atividades e operações florestais, devem ser considerados os fatores que influenciam o crescimento da floresta e, dessa forma, anteceder o conhecimento dos rendimentos obtidos em termos de produção volumétrica, uma vez que boa parte das atividades que envolve o manejo florestal são decididas pela prognose de produções futuras e informações correntes (CAMPOS; LEITE, 2017). Nesse sentido, é necessário compreender os processos de crescimento e produção para o gerenciamento de florestas.

Campos e Leite (2017) definem crescimento como a mudança ocorrida na forma e no tamanho do tronco da árvore devido à deposição de camadas no lenho ao longo do material lenhoso, enquanto a produção é a quantidade total acumulada de alguma variável, por exemplo, o volume, em um determinado tempo.

Os modelos de crescimento e produção permitem gerenciar diversas atividades do setor florestal, como prescrever regimes de manejo adequados para cada espécie em determinados sítios (ABREU et al., 2002), obter a idade ótima de colheita (GUIMARÃES, 1994), e análise da viabilidade econômica de projetos florestais (SCOLFORO, 2006). Tais modelos podem ser classificados como: modelos em nível de povoamento, modelos por classe diamétrica e modelos de árvores individuais (DAVIS; JOHNSON, 2000).

Modelos em nível de povoamento não explicam diretamente a variação do tamanho da árvore dentro do povoamento. Por isso, são adequados para situações em que objetivo é apenas a obtenção do volume de madeira ou biomassa por unidade de área, como ocorre nas empresas de fabricação de carvão e celulose (CAMPOS; LEITE, 2017). Nestes modelos, a variável de interesse é obtida com base nas informações relacionadas ao povoamento, como: idade, sítio e área basal (CARVALHO, 2010).

Os principais modelos de crescimento e produção em nível de povoamento são (CAMPOS; LEITE, 2017):

### Schumacher

$$Y = e^{\beta_0 + \beta_1 \left( \frac{1}{I} \right)} + \varepsilon \quad (1)$$

### Logístico

$$Y = \frac{\beta_0}{1 + \beta_1 e^{-\beta_2 I}} + \varepsilon \quad (2)$$

### Gompertz

$$V = \beta_0 + \beta_1 e^{-e^{\beta_1 - \beta_2 I}} + \varepsilon \quad (3)$$

### Richards

$$Y = \frac{\beta_0}{\left( 1 + e^{\beta_1 - \beta_2 I} \right)^{\frac{1}{\beta_3}}} + \varepsilon \quad (4)$$

### Clutter

$$\ln(V_2) = \beta_0 + \beta_1 \left( \frac{1}{I_2} \right) + \beta_2 S + \beta_3 \ln B_2 + \ln \varepsilon \quad (5)$$

$$\ln(B_2) = \ln B_1 \left( \frac{I_1}{I_2} \right) + \alpha_0 \left( 1 - \frac{I_1}{I_2} \right) + \alpha_1 \left( 1 - \frac{I_1}{I_2} \right) S + \ln \varepsilon \quad (6)$$

em que:  $\beta_0$ ,  $\beta_1$ ,  $\beta_3$ ,  $\alpha_0$  e  $\alpha_1$  = parâmetros a serem ajustados;  $I_1$  e  $I_2$  = idade atual e futura;  $B_1$  e  $B_2$  = área basal atual e futura;  $V_2$  = volume futuro;  $S$  = índice de sítio e  $\varepsilon$  = erro aleatório

O modelo mais difundido dentre os de povoamento total é o de Clutter, o qual utiliza-se de duas equações, uma para estimar a área basal em uma idade futura e outra para calcular o volume em uma idade futura. O modelo de Clutter é considerado compatível, consistente e explícito. Compatível porque a integral da equação de crescimento fornece a equação de produção e sua derivada resulta na equação de crescimento. Consistente porque as estimativas podem ser obtidas projetando-se a área basal ano a ano, ou diretamente de um ano para qualquer outro, sempre irá

fornecer os mesmos valores, e explicito porque o volume é calculado diretamente (CAMPOS; LEITE, 2017).

Os modelos de crescimento e produção do tipo distribuição por classe de diâmetro possuem maior detalhamento do que o de povoamento total. Isso facilita o planejamento da produção, como por exemplo, otimizar os rendimentos do povoamento por meio da utilização de multiprodutos (SCOLFORO, 2006). Tais modelos estimam o número de indivíduos por hectare por classe de diâmetro nas idades presente e futura. Para isso, utiliza-se de funções probabilísticas, como por exemplo: Sb Jonson, Lognormal, Logística Generalizada, Gompertz, Hiperbólica, Gumbel, Nakagami, Frechet, Dagum, Pearson, Burr, Gama e a comumente usada função Weibull (CAMPOS; LEITE, 2017).

Modelos de árvores individual consideram o crescimento de cada árvore simulando incrementos em altura, diâmetro, competição e a probabilidade de mortalidade de cada árvore (DAVIS; JOHNSON, 2000; CAMPOS; LEITE, 2017). Estes modelos são mais detalhados que os modelos em nível de povoamento e os de distribuição diamétrica, porém, são mais complexos, principalmente porque são necessários submodelos para estimar a mortalidade e a competição (HASENAUER, 2000), sendo que a competição é considerada uma variável de difícil determinação (ZHANG et al., 1997).

### 3.2 *Random Forest*

A *Random Forest* (RF) ou Florestas Aleatórias é uma técnica de aprendizagem de máquinas que pode ser usada para regressão e classificação (BREIMAN, 2001). A RF gera um conjunto de árvores de decisão e o resultado predito é a média de todas as árvores (TOOKE; COOPS; WEBSTER, 2014), para os casos de regressão. Nos problemas de classificação, o resultado obtido é dado por um sistema de votação, em que, a nova classe predita é aquela que obteve o maior número de votos (BREIMAN, 2001).

As árvores de decisão consistem em dividir de forma iterativa o banco de dados (nó pai) em subconjuntos (nós filhos) (ZAMO et al. 2014). A divisão dos dados é

calculada de acordo com algum critério (SHAIKHINA et al., 2017). Normalmente, é feito por um limiar de corte das variáveis independentes, afim de diminuir a variância da variável dependente nos nós filhos formados (JAMES et al., 2013; ZAMO et al., 2014).

Para realizar a divisão na árvore de decisão, seleciona-se a variável independente  $x_j$  e o ponto de corte  $s$ , de modo que, ao dividir o espaço da variável dependente, leva à maior redução possível da variância dos nós filhos, ou  $R_1$  (7) e  $R_2$  (8).

$$R_1(j, s) = \{x / x_j < s\} \quad (7)$$

$$R_2(j, s) = \{x / x_j \geq s\} \quad (8)$$

A notação  $\{x / x_j < s\}$ , significa o ponto de corte do espaço da variável independente em que  $x_j$  assume valor menor que  $s$ , considerando todas as variáveis independentes  $x_1, \dots, x_j$  e todos os valores possíveis para  $s$ . Dessa forma, busca-se o valor de  $j$  e  $s$  que minimize a Equação 9 (JAMES et al., 2013).

$$Erro = \sum_{i: x_i \in R_1(j, s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j, s)} (y_i - \hat{y}_{R_2})^2 \quad (9)$$

em que:  $y_i$  =  $i$ -ésimo valor da variável dependente;  $\hat{y}_{R_1}$  = valores estimados na região 1 ou nó filho;  $\hat{y}_{R_2}$  = valores estimados na região 2 ou nó filho.

Esse processo ocorre iterativamente até que um critério de parada seja atingido, que pode ser o número final de observações. O grupo final é chamado de folha e o valor previsto é a média dos valores pertencentes as folhas das árvores (ZAMO et al., 2014).

Alguns dos algoritmos usados pela RF são o *bagging aggregation* e o *Classification and Regression Tree* (CART). O CART é o algoritmo que faz o particionamento recursivo da árvore de decisão. O *bagging aggregation* combina várias árvores de decisão usando uma amostragem *bootstrap* (BREIMAN, 2001). As

amostras são formadas de maneira aleatorizada, com reposição e sem podas a partir do banco de dados original (HASTIE et al., 2009; YU et al., 2011). Este conjunto é indicado por  $\hat{f}^1(x), \hat{f}^2(x), \dots, \hat{f}^B(x)$  e pode ser ajustado por meio da Equação 10.

$$f_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x) \quad (10)$$

em que:  $B$  = o número de árvores;  $f_{bag}(x)$  = conjunto de árvores;  $\hat{f}^b$  = árvore de decisão.

A variância para os conjuntos de dados previstos no *baagging* ( $f_{bag}(x)$ ) é obtida por meio da Equação 11:

$$Var(f_{bag}(x)) = \rho \sigma^2 + \frac{1-\rho}{B} \sigma^2 \quad (11)$$

em que:  $\rho$  = correlação;  $\sigma^2$  = variância para o conjunto de árvores.

O *bootstrap* é uma abordagem estatística usada para quantificar a incerteza associada a um determinado estimador (JAMES et al., 2013). Esse método gera várias amostras com reposição (CANO et al., 2017). Dessa forma, uma amostra já escolhida é colocada de volta no conjunto original de dados e alguns registros podem ser escolhidos novamente. De acordo com Grabczewski (2014), pode ser amostrado em torno de 63,2% dos registros nos dados originais. Consegue-se fazer essa aproximação, porque a probabilidade de um registro ser escolhido a partir de uma amostra de *bootstrap* é de  $1 - (1 - 1/N)^N = 0,632$ , sendo  $N$  o número de registros. Tal procedimento é repetido  $n$  vezes para gerar  $n$  amostras *bootstrap*. Os registros que não são inseridos nas amostras *bootstrap* poderão fazer parte do conjunto de teste (TAN; STEINBACH; KUMAR, 2006). Estas amostras, são chamadas *out-of-bag* (oob) e podem ser utilizadas para estimar o erro de classificação ou regressão das RF (YU et al., 2011; JAMES et al., 2013). A soma do erro oob e a importância média de todas as árvores determinam as variáveis de maior importância, total e relativa, que serão selecionadas (CANO et al., 2017), ou seja, a RF classifica as variáveis de acordo com uma contribuição na previsão da variável dependente.

A vantagem do uso da RF comparada aos vários métodos de aprendizagem de máquinas, como exemplo, as máquinas de vetores de suporte e as redes neurais artificiais, são as informações obtidas sobre as variáveis que melhor colaboram para previsão do modelo (ARCHER; KIMES, 2008).

A importância das variáveis independentes para previsão do modelo pode ser calculada usando duas medidas diferentes de importância (CANO et al., 2017). A primeira medida de importância é o erro médio quadrático (MSE) para regressão, em que é calculada com a diferença entre o MSE do nó pai e o MSE total dos nós filhos após a divisão.

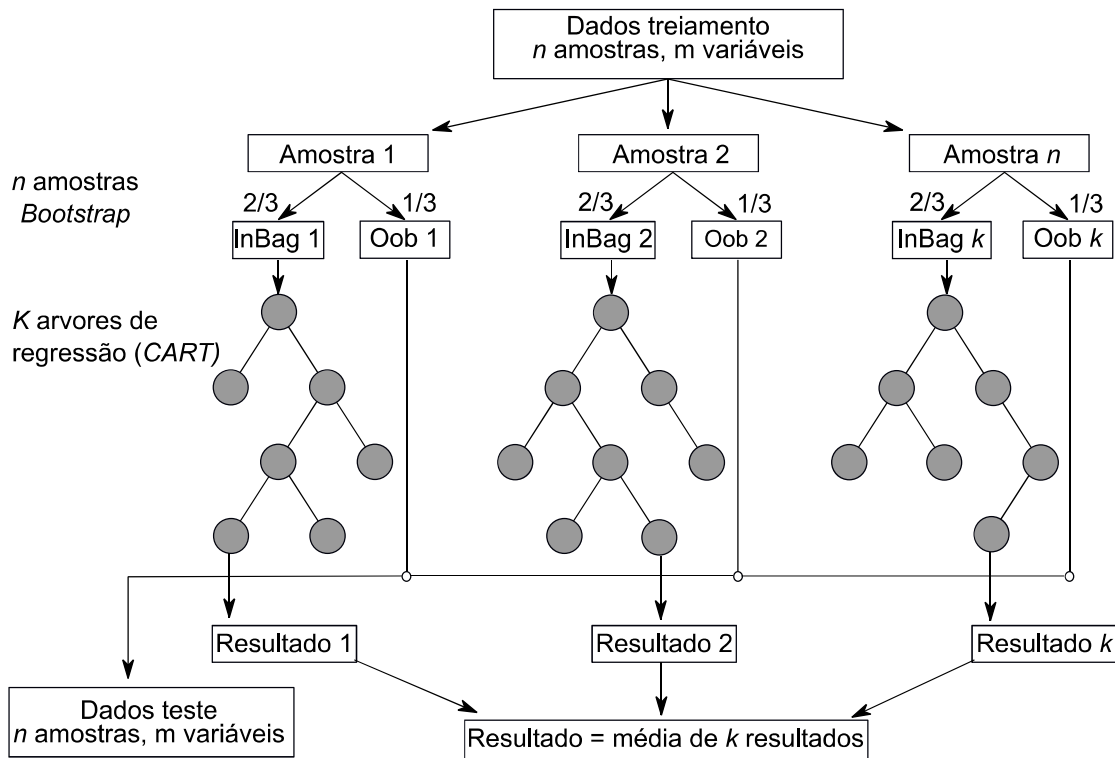
A segunda medida de importância é o índice Gini para classificação, que é o critério usado para selecionar a melhor divisão no desenvolvimento das árvores. Esta é baseada no grau de impureza dos nós filhos, que são formados depois da divisão (TAN; STEINBACH; KUMAR, 2006). O índice Gini mede o erro de classificação no nó  $t$  e pode ser calculado de acordo com a Equação 12:

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i/t)]^2 \quad (12)$$

em que:  $c$  = número de classes;  $p$  = probabilidade estimada na classe  $i$  para as instâncias que atinge o nó  $t$  (TAN; STEINBACH; KUMAR, 2006). Na Figura 1 é demonstrado o mecanismo de obtenção das repostas numéricas na RF.



Figura 1 – Mecanismo de funcionamento da *Random Forest*



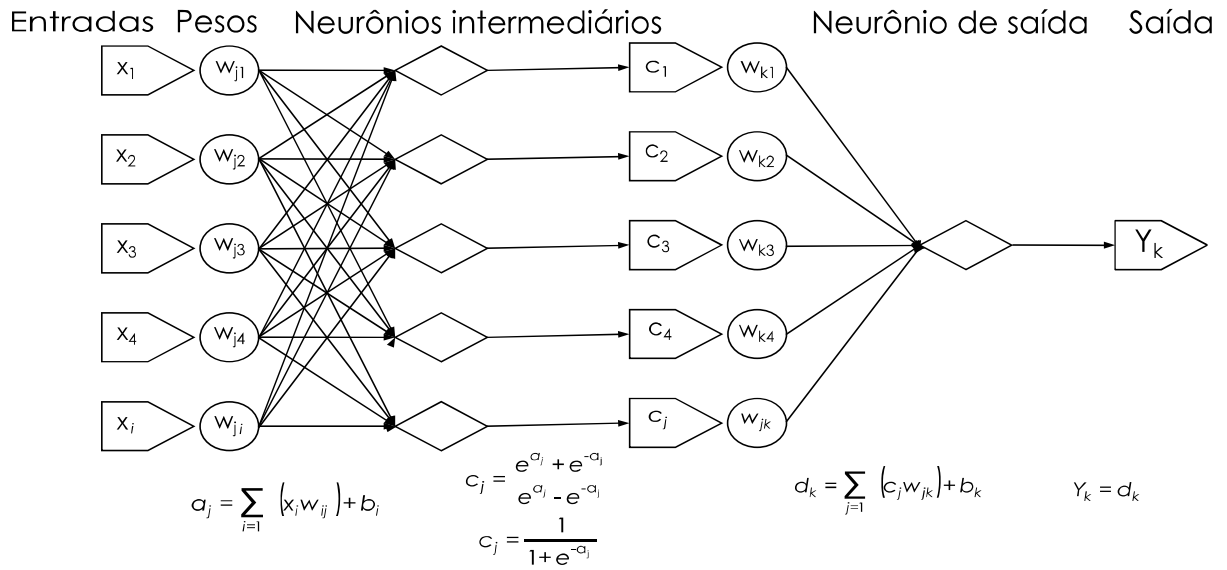
Fonte: adaptado de Ibrahim e Khatib (2017).

### 3.3 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas computacionais com unidades de processamento simples que calculam certos tipos de funções matemáticas, normalmente não-lineares (BRAGA; CARVALHO; LUDERMIR, 2007), como as funções logarítmicas, tangente hiperbólica e gaussianas. O poder computacional da rede é extraído da sua estrutura paralela e distribuída. Com isso, possui habilidade de aprender e generalizar o conhecimento adquirido, tornando-a capaz de solucionar problemas complexos (HAYKIN, 2001).

O conhecimento da RNA do tipo *feedforward* é armazenado nas conexões entre unidades de processamento, chamados de pesos sinápticos, que são obtidos por um processo de adaptação ou aprendizagem (BRAGA; CARVALHO; LUDERMIR, 2007). A estrutura de uma RNA do tipo *feedforward* é apresentada na Figura 2.

Figura 2 – Rede neural artificial.



Fonte: adaptado de Simões e Shaw (2007).

A camada de entrada é responsável por receber os sinais de entrada  $x_i$ . Cada entrada ( $x_i$ ) é multiplicada por pesos sinápticos ( $w_{ij}$ ) na camada intermédia, e posteriormente ocorre a soma das entradas ponderadas, incluindo o bias ( $b_i$ ), que serve para aumentar ou diminuir a entrada líquida da função de ativação (Eq.13).

$$a_j = \sum_{i=1} (x_i w_{ij}) + b_i \quad (13)$$

O somatório das entradas ponderadas ( $a_j$ ) é transformado por uma função de ativação normalmente não-linear, por exemplo, a função logística (Eq.14) ou a tangente hiperbólica (Eq.15). O intervalo padronizado destas funções é de 0 a 1 ou de -1 a 1, respectivamente (HAYKIN, 2001).

$$c_j = \frac{1}{1 + e^{-a_j}} \quad (14)$$

$$c_j = \frac{e^{a_j} + e^{-a_j}}{e^{a_j} - e^{-a_j}} \quad (15)$$

Depois que passa pela camada intermediária o valor de  $c_j$  se torna entrada do neurônio da camada de saída. Cada valor de  $c_j$  é multiplicado por pesos sinápticos ( $\omega_{jk}$ ), e posteriormente ocorre a soma das entradas ponderadas junto com o bias (Eq.16).

$$Y_k = d_k = \sum (c_j w_{jk} + b_k) \quad (16)$$

em que  $\hat{Y}_k$  = saída estimada;  $b_k$  = bias.

O passo seguinte é calcular o erro e corrigir os pesos com algoritmo de aprendizagem. Os algoritmos mais comuns são: *Backpropagation*, *resilient-propagation* e o Levenberg-Marquardt.

### 3.4 Lógica Fuzzy

A lógica *fuzzy* é uma extensão da lógica clássica booleana que foi proposta por Zadeh (1965), e permite que uma função de pertinência retorne valores em intervalos pertencentes entre 0 a 1, isto, possibilita uma maior flexibilidade comparada à lógica booleana (SILVA, 2006). De acordo com Jang (1993), os sistemas *fuzzy*, geralmente possuem os seguintes componentes: fuzzificador, conjunto de regras, funções de pertinência, inferência *fuzzy* e defuzzificador.

A fuzzificação associa a cada valor de entrada do modelo *fuzzy* um valor no intervalo [0,1], que corresponde ao grau de sua pertinência (ARAÚJO, 2004). Nesse estágio ocorre a ativação das regras de inferências utilizadas para associar as funções de pertinência das entradas com as funções de pertinência da saída (SHAW; SIMÕES, 2007).

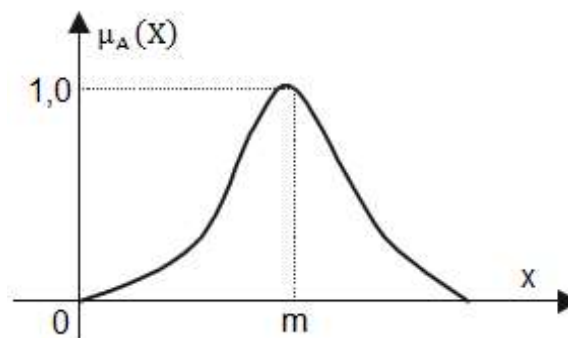
A regra *fuzzy* é composta por conjunto de sentenças condicionais do tipo “se....então”, onde irá formar o conhecimento especialista do sistema. Operadores lógicos conectam as variáveis de entrada entre si, e são associadas aos conjuntos

*fuzzy*, em todo universo de discurso, por meio das funções de pertinência (ARAÚJO, 2004). Os operadores lógicos podem ser "END", "OR" e "NOT" (MATHWORK, 2017).

As funções de pertinência criam um modelo que transforma as variáveis numéricas em correspondentes linguísticos (MASSELLI, 2009). Isso acontece da seguinte forma: um conjunto *fuzzy* A do universo do discurso X é descrito por uma função de pertinência  $\mu_A$  que associa a cada elemento x de X um número  $\mu_A(x)$  no intervalo [0,1], o qual representa o grau de pertinência de x em A (ARAÚJO, 2004). Desse modo, torna-se possível um elemento pertencer a mais de um conjunto *fuzzy* com diferentes graus de pertinência. As funções mais comuns são do tipo gaussiano, triangular e trapezoidal (MATHWORK, 2017).

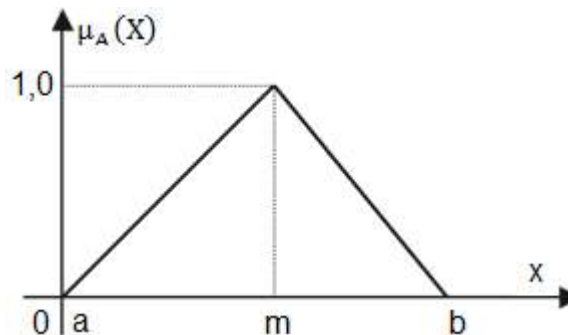
Nas Figuras 3, 4 e 5 e nas Equações 17, 18 e 19 estão descritas as funções gaussianas, triangulares e trapezoidais, respectivamente.

Figura 3 – Função gaussiana



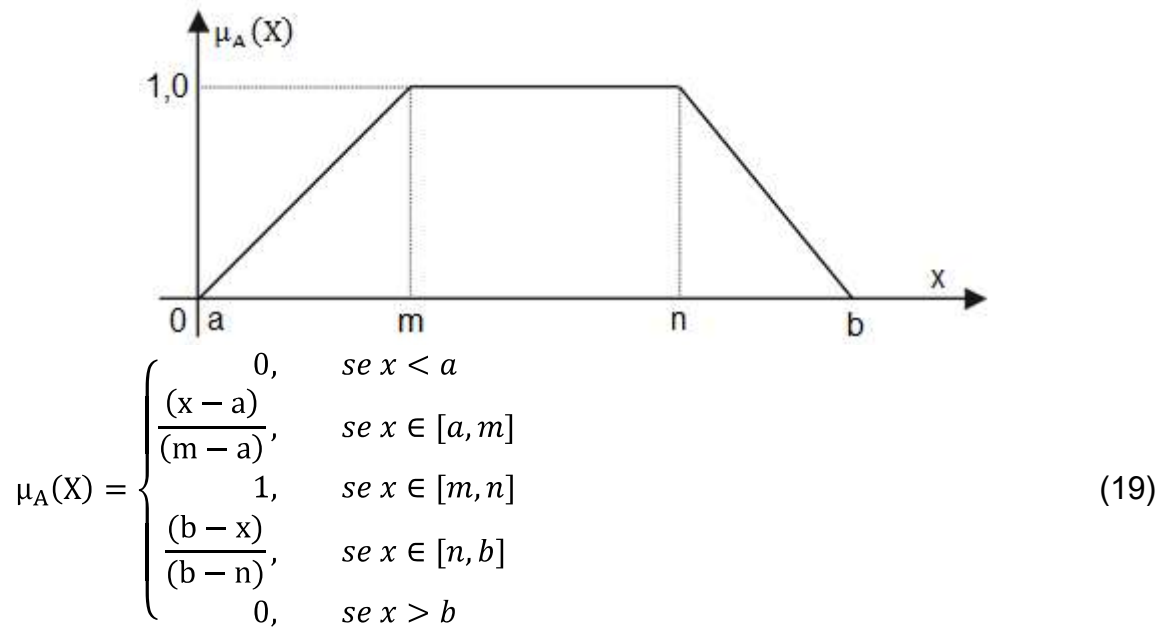
$$\mu_A(X) = \begin{cases} 0, & \text{se } x \text{ está fora do domínio} \\ e^{\frac{-(x-m)^2}{2s^2}} & \end{cases} \quad (17)$$

Figura 4 – Função triangular



$$\mu_A(X) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{(x-a)}{(m-a)}, & \text{se } x \in [a, m] \\ \frac{(b-x)}{(b-m)}, & \text{se } x \in [m, b] \\ 0, & \text{se } x \geq b \end{cases} \quad (18)$$

Figura 5 – Função trapezoidal



A inferência *fuzzy* estabelece a relação entre as regras, reunindo o conhecimento necessário para uma determinada conclusão. A saída do mecanismo de inferência *fuzzy* corresponde os termos linguísticos com seus devidos graus de pertinência, necessários para o processo de defuzzificação (MASSELLI, 2009). De acordo com Araújo (2004), no processo de defuzzificação acontece a transformação das variáveis linguísticas, gerada pela inferência, em um valor numérico. Existem diversos métodos para defuzzificação, sendo os mais comuns o método do centróide e os que são baseados nos valores máximos.

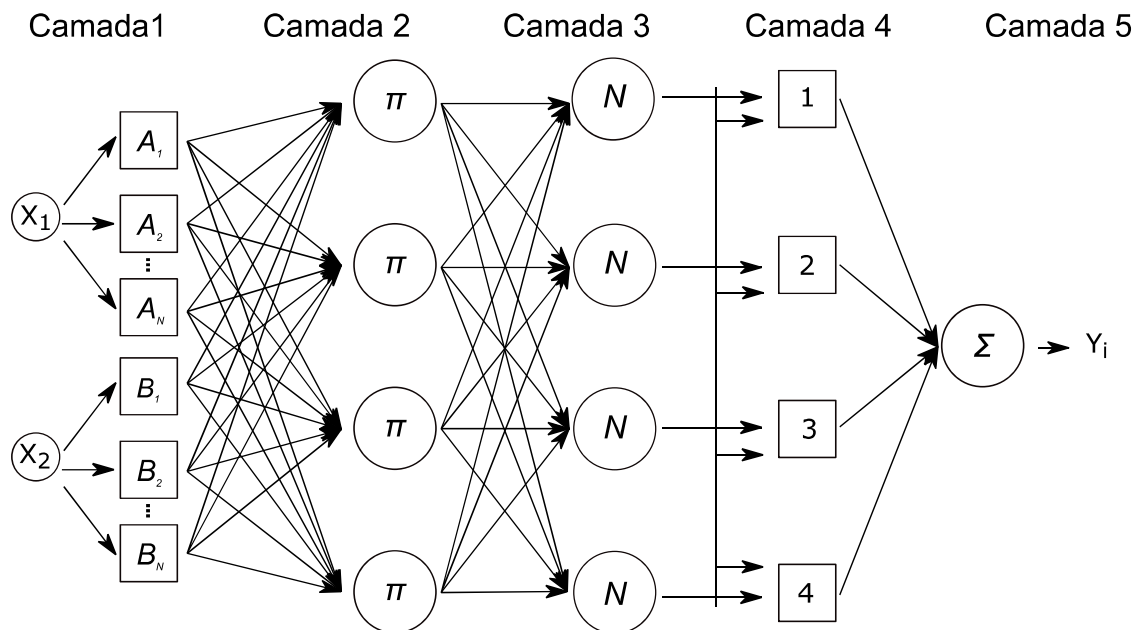
### 3.5 Sistema *Neuro-Fuzzy*

O SNF, pode ser definido como um sistema de inferência *fuzzy* treinado por uma RNA. Esta combinação reúne a capacidade das RNA em reconhecimento de

padrões, classificação, estimação de funções não-lineares, robustez e habilidade de generalização (BRAGA; CARVALHO; LUDERMIR, 2007), com a facilidade do entendimento e interpretação do problema proposta pela lógica *fuzzy*. Desse modo, esta combinação possibilita fazer inferências de situações da vida real de forma mais parecida com o raciocínio humano (PAL; MITRA, 1992). Os sistemas *neuro-fuzzy* tem apresenta-se uma técnica com alta capacidade de modelar sistemas da vida real (SIMINSKI, 2017).

O *Adaptive Neuro-Fuzzy Inference System* (ANFIS), proposto por Jang (1993), é uma rede neural com a função básica de implementar um sistema de inferência *fuzzy* por meio de uma arquitetura paralela distribuída de uma RNA, de maneira que os algoritmos de aprendizagem possam ser utilizados para ajuste do sistema de inferência *fuzzy*. O ajuste dos parâmetros associados com as funções de pertinência é efetuado utilizando o algoritmo de aprendizagem *backpropagation*, ou com uma combinação com o algoritmo dos mínimos quadrados. Esta estrutura é do tipo Takagi-Sugeno (TAKAGI; SUGENO, 1985) e apresentando funções lineares ou constantes nos consequentes das regras que formam o sistema *fuzzy* (MATHWORKS, 2017). Na Figura 6, está demonstrado a estrutura de sistema *neuro-fuzzy* do tipo ANFIS com duas entradas e uma saída.

Figura 6 – Estrutura do sistema *neuro-fuzzy*.



Fonte: adaptado de Fallahian e Seyedpoor (2010).

Na primeira camada ocorre o processo de fuzzificação, em que o conjunto universo dos números reais são mapeados para o domínio *fuzzy* (MATHUR; GLESK; BUIS, 2016). Cada nó desta camada cria um grau de associação ( $\mu_i$ ) para as variáveis de entrada ( $x_1$  e  $x_2$ ) que varia de 0 a 1 utilizando funções de pertinência ( $A_j$ ,  $B_k$ ), por exemplo, a função gaussiana, descrita pelas Equações 20 e 21 (BILGEHAN, 2011).

$$O_{1,j} = \mu_{A_j}(x_1) = \exp\left[-\left(\frac{x_1 - c_j}{a_j}\right)\right] \quad (20)$$

$$O_{1,k} = \mu_{B_k}(x_2) = \exp\left[-\left(\frac{x_2 - c_k}{a_k}\right)\right] \quad (21)$$

em que:  $c_j$  e  $a_j$  = parâmetros dos antecedentes a serem estimados pelo gradiente descendente.

Na camada dois, os nós são fixos são denominados como  $O_{2,i}$ . A saída de cada nó é o produto de todos os sinais de entrada descrito pela Equação 22 (AKBARZADEH et al., 2014; MATHUR; GLESK; BUIS, 2016).

$$O_{2,i} = \omega_i = \mu_{A_k}(x_1)\mu_{B_k}(x_2) \quad (22)$$

Na camada três os valores de  $\omega_i$  são normalizados no nó  $O_{3,i}$  por meio da Equação 23.

$$O_{3,i} = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2} \quad (23)$$

Na camada quatro acontece a defuzzificação, que é a transformação dos valores *fuzzy* para um número real (SIMÕES; SHAW, 2007), como descrito pela Equação 24.



$$O_{4,i} = \bar{\omega}_i f_i = \bar{\omega}_i (a_i x_1 + b_i x_2 + c_i) \quad (24)$$

As constantes  $a_i$ ,  $b_i$  e  $c_i$  são parâmetros dos consequentes e podem ser ajustados por análise de regressão linear utilizando o método de mínimos quadrados.

A última camada (5) é a soma de  $\bar{\omega}_i f_i$  que irá fornecer a saída final por meio da Equação 25.

$$O_{5,i} = \sum_j \bar{\omega}_j f_j = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i} \quad (25)$$

A saída do Sistema *Neuro-Fuzzy* é modificada de acordo com os parâmetros dos antecedentes e consequentes. Estes parâmetros podem ser ajustados com uma abordagem híbrida, em que os parâmetros dos antecedentes são ajustados pelo método do gradiente descendente e dos consequentes pelo método dos mínimos quadrados (AKBARZADEH et al., 2014; MATHUR; GLESK; BUIS, 2016).

O processo de treinamento, de acordo com Chaves (2013), é constituído por duas etapas. Na primeira etapa ou *forward*, os parâmetros dos antecedentes são fixos e os parâmetros dos consequentes são ajustados via algoritmo dos mínimos quadrados. Na segunda etapa ou *backward*, os sinais dos erros são retropropagados (*backpropagation*), e os parâmetros dos antecedentes são atualizados pelo método do gradiente descendente.

O antecedente divide uma região no espaço de entrada enquanto que o consequente descreve o comportamento desta região por meio de uma equação linear. Essa divisão é necessária para gerar as regras do sistema de inferência fuzzy e pode ser feita pelos algoritmos *Grid Partitioning* (GP), *Subtractive Clustering* (SC) e *Fuzzy-C-Means* (FCM).

### 3.6 Técnicas de inteligência artificial aplicadas na modelagem do crescimento e produção florestal

Castro et. (2013), trabalharam com redes neurais artificiais do tipo *multilayer perceptron* para prognosticar o crescimento de árvore individual. A RNA foi superior na projeção do crescimento e na estimação da probabilidade de mortalidade das árvores, comparada aos resultados obtidos por Martins (2011), usando modelos de regressão na mesma área.

As RNA foram usadas para modelar o crescimento do volume de árvore individual em diferentes cenários de mudança climática no Canadá (ASHRAF et al., 2015). As entradas usadas para o treinamento das RNA foram os dados de inventários florestais, radiação solar incidente, índices de nutrientes no solo e teor de água no solo. Foram feitas projeções sucessivas, em intervalos de 5 anos até 90 anos. Os valores de coeficiente de correlação obtidos variaram de 0,90 a 0,95.

Reis et al. (2016), usaram RNA para realizar a prognose do diâmetro de árvores nativas na floresta amazônica. A RNA conseguiu modelar a tendência de crescimento por um período de 33 anos com valores de RMSE (raiz quadrada do erro médio) abaixo de 11% e com valores de coeficiente de correlação acima de 0,99.

Reis (2017), aplicou RNA para modelar a tendência de recrutamento ao longo do tempo após a colheita de madeira em florestas tropicais. Dentre as variáveis utilizadas como entrada, as que mais influenciaram na estimação do recrutamento, foram o período de tempo entre as medições e o período de tempo após a colheita. A RNA foi adequada para predição do recrutamento de florestas tropicais.

Reis et al. (2018), trabalharam com RNA do tipo *multilayer perceptron* para prognosticar a sobrevivência e a mortalidade de árvores individuais em uma floresta na Amazônia, após a colheita florestal. A RNA foi eficiente para prognosticar a sobrevivência com taxas de acerto acima de 99% e coeficiente kappa superior a 55%. Ainda conforme estes autores, a RNA apresentou excelentes resultados para prognosticar a distribuição de diâmetros ao final do ciclo de corte, porém, não foram eficientes para prognosticar a mortalidade apresentando taxas de acertos abaixo de 6%.

Vieira et al. (2018), trabalharam com sistemas *neuro-fuzzy* e redes neurais artificiais para modelar o crescimento de árvore individual. Esses autores, utilizaram

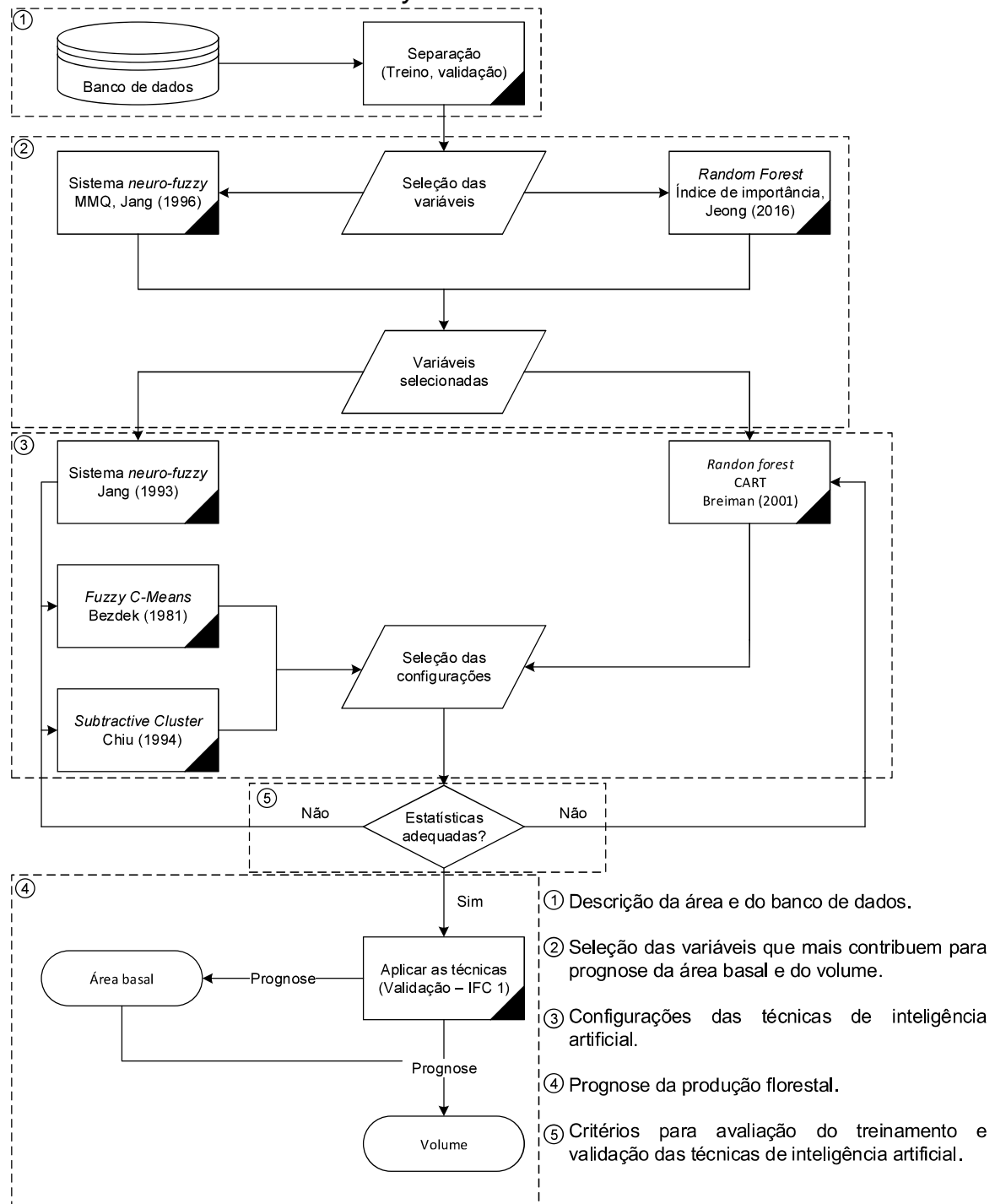
variáveis como idade atual e futura, material genético, sítio, índice de competição, diâmetro atual e altura atual como entrada das técnicas para modelar o crescimento do DAP e da altura de árvores de eucalipto e concluíram que as técnicas apresentaram maior acurácia para projetar o DAP e a altura das árvores, comparadas aos modelos de regressão, sendo, portanto, indicadas para estes fins.

Reis (2018), utilizou autômatos celulares e RNA para projetar a distribuição diamétrica em uma floresta colhida em área de manejo florestal sustentável na Amazônia. A distribuição diamétrica projetada ao longo do tempo não diferiu significativamente a 5% de probabilidade da distribuição observada ao longo de 30 anos. Além disso, verificou que a projeção segue a tendência em todo o ciclo de corte dos valores observados, sem apresentar desvios da forma de exponencial negativa da floresta. A RNA que apresentou as melhores estatísticas para generalização, (RMSE de 12,95% e coeficientes de correlação de 0,9976), continham 10 neurônios na camada oculta com função de ativação logística e um neurônio na camada de saída com a função exponencial.

## 4 MATERIAL E MÉTODOS

As etapas realizadas para projetar a área basal e o volume de povoamentos de eucalipto com o SNF e a RF estão demonstradas na Figura 7.

Figura 7 – Fluxograma metodológico para realização da prognose da produção florestal com Sistema *Neuro-Fuzzy* e *Random Forest*.



#### 4.1 Descrição da área de estudo e do banco de dados

Os dados utilizados no presente estudo foram obtidos a partir inventários florestais contínuos (IFC's), realizados em povoamentos clonais de eucalipto. A área tem aproximadamente 2000 ha e está localizada no sul da Bahia.

O número de medições variou de 4 a 10, em um total de 276 parcelas permanentes, com área de 214 m<sup>2</sup>. As variáveis numéricas empregadas neste estudo foram idade (meses), área basal (m<sup>2</sup>.ha<sup>-1</sup>), volume (m<sup>3</sup>.ha<sup>-1</sup>), sendo que o volume considerado foi medido até o diâmetro mínimo de 4 cm com casca. As variáveis categóricas foram: clone (C), espaçamento (E), tipo de solo (S), relevo (R) e textura do solo (T), a quantidade de classes dessas variáveis foi 8, 6, 15, 3 e 7, respectivamente.

As parcelas com todas as medições foram separadas em 70% para treinamento e 30% para validação do SNF e da RF. De acordo com Simões e Shaw (2007), o conjunto de dados de treinamento deve abranger a faixa total (amplitude das variáveis quantitativas e classes das variáveis qualitativas) de entradas e saídas dos dados de validação. Seguindo esta recomendação foram feitas análises descritivas dos dados são apresentadas na Tabela 1.

Tabela 1 – Estatísticas descritivas para o conjunto de treinamento e validação do Sistema *Neuro-Fuzzy* e *Random Forest*

Variável	Treinamento				Validação			
	Mín	Máx	Méd	DP	Mín	Máx	Méd	DP
Área basal (m <sup>2</sup> ha <sup>-1</sup> )	1,1	48,8	20,5	±6,0	2,5	33,5	20,6	±5,5
Idade (meses)	21,3	136,7	70,9	±25,3	21,3	136,4	71,3	±25,4
Volume (m <sup>3</sup> ha <sup>-1</sup> )	12,1	1111,7	438,9	±193,5	31,1	887,7	427,6	±180,3

em que: Mín = mínimo; Máx = máximo; DP = desvio padrão.

## 4.2 Seleção das variáveis que mais contribuem para prognose da área basal e do volume

Como o banco de dados para projetar a área basal e volume dispõe de dezenas de entradas potenciais para construção do SNF e da RF, foi feita uma seleção para priorizar apenas as entradas que mais influenciaram nas estimativas. Utilizou-se as próprias técnicas para selecionar as variáveis.

A seleção com o SNF foi com base na metodologia proposta por Jang (1996), que consiste em treinar várias combinações de variáveis de entrada em apenas uma época com o algoritmo de mínimos quadrados e escolher a que apresentar o menor RMSE (raiz quadrada do erro médio). Posteriormente, a combinação selecionada foi treinada com mais épocas e configurações.

As variáveis de maior importância na RF foram escolhidas pela diminuição da impureza do nó, que é calculada com o erro quadrado médio após a divisão (JEONG et al., 2016).

## 4.3 Configurações das técnicas de inteligência artificial

### 4.3.1 Sistema *Neuro-fuzzy*

Foram utilizados os algoritmos de agrupamento *Fuzzy C-Means* (FCM) Bezdek (1981) e *Subtractive Cluster* (SC) Chiu (1994) para otimizar o número de regras geradas no treinamento do SNF. O agrupamento consiste em encontrar pontos de dados com maior semelhança em um mesmo cluster e menor semelhança entre dados de clusters diferentes (SU; ZHAO, 2017).

O FCM encontra os centros dos clusters, minimizando a distância de  $x_k$  e  $v_i$ , com a função objetivo representada pela Equação 26.

$$J = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m \|x_k - v_i\|^2 \quad (26)$$

$$v_i = \frac{\sum_{k=1}^n [\mu_i(x_k)]^2 x_k}{\sum_{k=1}^n [\mu_i(x_k)]^m} \quad (27)$$

$$\mu_{ik} = \frac{1}{\sum_{i=1}^c \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{1/m-1}} \quad (28)$$

em que:  $n$  = número de observações;  $c$  = número de clusters;  $x_k$  = k-ésima observação;  $v_i$  = i-ésimo centro do cluster;  $m$  = constante exponencial com valor maior que 1;  $\mu_{ik}$  = grau da função de associação e inicialmente ocorre de forma aleatória.

O algoritmo FCM precisa de três parâmetros: o número de clusters, valor do expoente e o número de iterações. Neste trabalho foram analisados o número de cluster entre 6 e 22, expoente de 1,115 variando em intervalos de 0,100 até 2,015 e com o número de iterações igual a 100.

O algoritmo SC considera que cada ponto dos dados tem potencial de ser o centro do cluster, que é calculado conforme a Equação 29:

$$P_i^* = \sum_{j=1}^n e^{-\alpha \|x_i - x_j\|^2} \quad (29)$$

$$\alpha = \frac{4}{r_a^2} \quad (30)$$

em que  $P_i^*$  = potencial do ponto  $i$  ser o primeiro centro de cluster;  $x_i$  = ponto candidato a ser o centro de cluster;  $r_a$  = o raio de influência do cluster.

Após os cálculos com a Equação 30, seleciona-se o ponto que apresentou o maior potencial para ser o primeiro centro (Equação 29). Os pontos que estão fora do alcance do raio de influência são removidos. Os próximos clusters são formados de acordo com a Equação 31, em que, os pontos próximos ao primeiro cluster tem seus potenciais reduzidos e não são escolhidos como centro. Esse processo ocorre iterativamente.



$$P_i \Leftarrow P_i - P_k^* e^{-\beta \|x_i - x_k^*\|^2} \quad (31)$$

$$\beta = \frac{4}{\eta r_a} \quad (32)$$

em que  $P_k^*$  = potencial dos pontos que já são centros dos clusters;  $x_k^*$  = centros dos cluters formados;  $r_a$  = raio de influência do cluster;  $\eta$  = fator *squash*.

Para fazer o agrupamento dos dados para estimar a produção da área basal e o volume utilizando o SC, foi preciso especificar quatro parâmetros:

1. Fator *squash*. Reduz a probabilidade de pontos periféricos serem considerados parte de um cluster. Os valores analisados variaram entre 0,95 e 1,55 em intervalos de 0,1.
2. Raio de influência. O raio é o alcance que o centro do cluster tem sob os pontos vizinhos. Pontos com distâncias menores que o alcance do raio são agrupados no mesmo cluster. Os valores analisados variaram entre 0,1 e 0,9 em intervalos de 0,05.
3. Taxa de aceitação. Considera apenas pontos com potenciais de serem centros do cluster. O seu valor foi de 0,5 (CHIU, 1994).
4. Taxa de rejeição. Rejeita os pontos com valores abaixo do potencial do primeiro cluster. O seu valor foi de 0,15 (CHIU, 1994).

O processamento dos dados com a SNF foi realizado com o *toolbox logic fuzzy* do Software Matlab R2016a. Este *toolbox* possui o módulo ANFIS (*Adaptive Neuro-Fuzzy Inference System*), proposto por JANG (1993).

As funções de associação dos clusters foram todas do tipo gaussianas e para saída a função linear. O algoritmo de treinamento foi o híbrido, que consiste, na combinação do método *backpropagation* com o método dos mínimos quadrados. O critério de parada para o treinamento foi o erro igual a zero ou a *early stoopping* (parar mais cedo). A *early stoopping* é utilizada para melhorar a generalização e evitar o *overfit* dos dados. O número de épocas de treinamento variou de 1 a 20.

#### 4.3.2 Random Forest

O processamento dos dados com a RF foi realizado com o *toolbox Statistics and Machine Learning* do Software Matlab R2016a. O mecanismo de treinamento da *random forest* consiste em criar várias árvores de regressão por meio do algoritmo CART. Estas árvores são geradas a partir de amostras *bootstrap*. Para projetar a área basal e o volume de árvores de eucalipto usando RF, foi preciso definir o número de árvores, número de variáveis escolhidas para começar a divisão do nó e o número de folhas para cada árvore formada. Neste trabalho, o número de árvores variou de 50 a 850, variando em intervalos de 50. O número de observações por folhas variou de 5 a 35 em intervalos de 5.

#### 4.4 Prognose da produção florestal

A área basal futura foi considerada como entrada para projeção do volume. Nesse caso também foram treinadas SNF e RF para projetar a área basal. Após o treinamento e validação das técnicas, as melhores configurações foram aplicadas no primeiro inventário florestal (IFC), realizado aos 24 meses, do conjunto de validação para realizar a projeção do volume e área basal.

#### 4.5 Critérios para avaliação do treinamento e validação das técnicas de inteligência artificial

A avaliação das estimativas geradas pelas técnicas SNF e RF, para treinamento, validação e prognose da área basal e volume foram calculadas com base nas seguintes estatísticas: coeficiente de correlação linear entre os valores

observados e estimados ( $r_{\hat{Y}}$ ), viés (V) e raiz quadrada do erro médio (RMSE) em porcentagem (Tabela 2).

Tabela 2 – Estatísticas para avaliar o desempenho das técnicas de inteligência artificial

Estatísticas	Fórmulas	Nº
Coeficiente de correlação	$r_{y\hat{y}} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \hat{Y}_m)}{\sqrt{\left[ \sum_{i=1}^n (Y_i - \bar{Y})^2 \right] \left[ \sum_{i=1}^n (\hat{Y}_i - \hat{Y}_m)^2 \right]}}$	(33)
Viés	$V(\%) = \frac{100}{\bar{Y}} \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{Y}_i}{n}$	(34)
Raiz quadrada do erro médio	$RMSE(\%) = \frac{100}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$	(35)

em que:  $Y_i$  = variável dependente observada,  $\hat{Y}_i$  = variável dependente estimada,  $\bar{Y}$  = média da variável dependente observada,  $\hat{Y}_m$  = média da variável dependente estimada e  $n$  = número de observações.

De forma a complementar a análise, foram elaborados gráficos relacionados aos valores observados e estimados pelas técnicas, gráficos de resíduos percentuais e histogramas de erros percentuais. O erro de cada observação foi calculado em porcentagem com a Equação 36.

$$E(\%) = \frac{Y_i - \hat{Y}_i}{Y_i} 100 \quad (36)$$

em que:  $E(\%)$  = erro de cada observação.

## 5 RESULTADOS E DISCUSSÃO

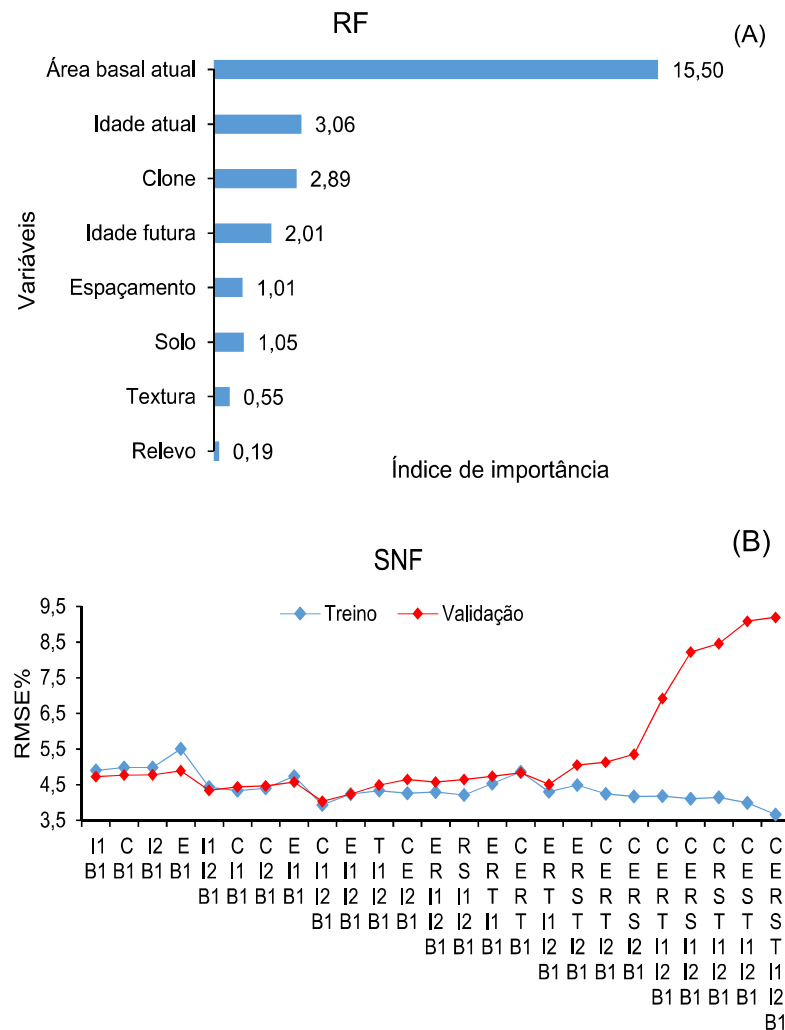
### 5.1 Área basal

#### 5.1.1 Seleção das variáveis de entrada para projetar a área basal

Em geral, todas as variáveis contribuem para a estimação da área basal futura (Figura 8A). Isto é verificado também no SNF (Figura 8B), em que o menor erro para o treinamento foi com todas as variáveis (RMSE = 3,66%). Porém, neste caso, o erro de validação (RMSE = 10,26%), foi superior às demais combinações. Isto é devido ao número de regras do SNF aumentar exponencialmente com a quantidade de entradas e consequentemente, os parâmetros lineares. Segundo Babuska e Verbruggen (2003), um modelo com muitos parâmetros pode aumentar o potencial da SNF de aproximar funções no treinamento, porém perde a capacidade de generalização. Mohammadi et al. (2016), afirma que para obter previsões com acurácia é essencial determinar as variáveis mais significativas no modelo do SNF.

Os resultados obtidos indicam que a área basal atual (B1) é a variável mais significativa para predição da área basal futura (B2) em ambas as técnicas, sendo a mais influente pelo método RF e a mais frequente nas melhores combinações do SNF. Além disso, o relevo (R) e a textura (T) são as variáveis menos relevantes (Figura 8A). Os resultados indicam que os valores de RMSE da validação aumentam e do treinamento diminui, conforme o aumento do número de variáveis de entradas (Figura 8B). Isto destaca a importância da seleção apropriada das variáveis para prever com certo nível de acurácia a área basal futura. Verificou-se ainda, que a combinação área basal atual, idade futura (I2), idade atual (I1) e o clone (C) forneceu o menor erro de validação, com valor de RMSE de 3,93% e treinamento de 4,03% (Figura 8B). Notou-se que estas são as variáveis que apresentaram o maior valor de índice de importância no gráfico do RF (Figura 8A).

Figura 8 – Variáveis mais influentes para prognose da área basal indicadas pela *Random Forest* (Figura 8A) e Sistema *Neuro-Fuzzy* (Figura 8B).



Verificou-se que a utilização de mais de cinco entradas pode não ser apropriada, pois gera aumento no erro de validação no SNF (Figura 8B). Nesse caso, a combinação mencionada acima é a mais indicada para atingir maior acurácia em previsões de área basal futura, com uma menor complexidade do modelo do SNF e da RF. Por isso, estas variáveis foram consideradas como entradas para o treinamento.

### 5.1.2 Configurações selecionadas das técnicas de inteligência artificial utilizadas para projetar a área basal

As combinações das configurações geraram 833 SNF treinados com o algoritmo SC, 833 com o algoritmo FCM e 833 com a RF. Destas, foram pré-selecionadas as duas configurações que obtiveram o menor erro de validação (Tabela 3).

Tabela 3 – Configurações pré-selecionadas do Sistema *Neuro-Suzzy* e *Random Forest*

Técnica	Algoritmo	N°	Parâmetros	
			Fator squash	Raio
Sistema <i>Neuro-fuzzy</i> (SNF)	SC	1	1,25	0,30
	SC	2	1,25	0,35
			N° de cluster	Expoente
	FCM	1	13	1,215
	FCM	2	11	1,415
			N° de árvores	N° de obs. por folha
<i>Random Forest</i> (RF)	CART	1	50	10
	CART	2	250	10

em que: SC = *Subtractive Cluster*; FCM = *Fuzzy-C-Means*; CART = *Classification and regression Tree*.

A escolha da melhor configuração para projetar a área basal futura baseou-se no erro de validação e na simplicidade do modelo gerado. Um modelo mais complexo pode aumentar a capacidade das técnicas em captar as relações de não linearidade dos dados. Porém, quanto mais complexo for o modelo, maior será a quantidade de regras geradas e consequente o tempo de processamento.

As configurações selecionadas para projetar a área basal foram denominadas de SNF-SC 1 (Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 1); SNF-SC 2 (Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 2); SNF-FCM 1 (Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 1); SNF-FCM 2 (Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 2); RF-CART 1 (*Random Forest* com algoritmo

*Classification and regression Tree* para configuração 1) e RF-CART 2 (*Random Forest* com algoritmo *Classification and regression Tree* para configuração 2).

### 5.1.3 Desempenhos das configurações das técnicas de inteligência artificiais selecionadas para prognose da área basal

Ao analisar as estatísticas utilizadas para verificar o desempenho do SNF e da RF (Tabela 4), para os dados de treinamento e validação, notou-se que, em geral, ambas as técnicas apresentaram bons resultados, com baixos valores de RMSE(%) e V(%) e um alto coeficiente de correlação. Tais estatísticas são similares às encontradas no trabalho de Binoti et al. (2014), para projetar a área basal com o uso de redes neurais artificiais. Os autores encontraram valores de RMSE que variaram de 4,62% a 5,20% para o treinamento e de 4,78% a 6,08% para generalização. Castro et al. (2013), utilizando equações simultâneas de Clutter, obteve erro padrão de 6,2% e coeficiente de correlação de 0,9674.

Tabela 4 – Estatísticas do treinamento e validação do Sistema *Neuro-Fuzzy* e *Random Forest* para projetar a área basal ( $\text{m}^2 \cdot \text{ha}^{-1}$ )

Técnica/configuração	Treino			Validação		
	$r_{Y\hat{Y}}$	V(%)	RMSE(%)	$r_{Y\hat{Y}}$	V(%)	RMSE(%)
SNF-SC 1	0,9916	0,00	3,80	0,9897	-0,29	3,90
SNF-SC 2	0,9900	0,00	4,13	0,9885	-0,19	4,09
SNF-FCM 1	0,9924	0,00	3,61	0,9909	-0,16	3,65
SNF-FCM 2	0,9940	0,00	3,20	0,9907	-0,13	3,69
RF-CART 1	0,9897	0,03	4,51	0,9857	-0,13	4,58
RF-CART 2	0,9893	0,04	4,68	0,9857	-0,22	4,61

em que: RMSE(%) = raiz do erro médio quadrático em porcentagem; v(%) = viés em porcentagem;  $r_{Y\hat{Y}}$  = coeficiente de correlação; SNF-SC 1 = Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 1; SNF-SC 2 = Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 2; SNF-FCM 1 = Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 1; SNF-FCM 2 = Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 2; RF-CART 1 = *Random Forest* com algoritmo *Classification and regression Tree* para configuração 1; RF-CART 2 = *Random Forest* com algoritmo *Classification and regression Tree* para configuração 2.

As estimativas obtidas pelas técnicas analisadas não foram viesadas na fase de treinamento para estimar a área basal futura. Já no conjunto de validação, o viés variou de -0,13 a -0,29. Castro et al. (2015), encontraram viés variando de -0,1139 a

-0,1145 em seus estudos sobre dois métodos para quantificar o índice de local empregado na modelagem do crescimento e produção.

Pode ser observado no conjunto de treinamento e validação (Figuras 9 e 10), que a maioria dos dados situam-se em torno da linha de 45° para ambas as técnicas, o que confirma que os dados observados de área basal estão fortemente relacionados com os valores estimados, como demonstrados pelos coeficientes de correlação na Tabela 4.

Figura 9 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) *versus* estimados no treinamento.

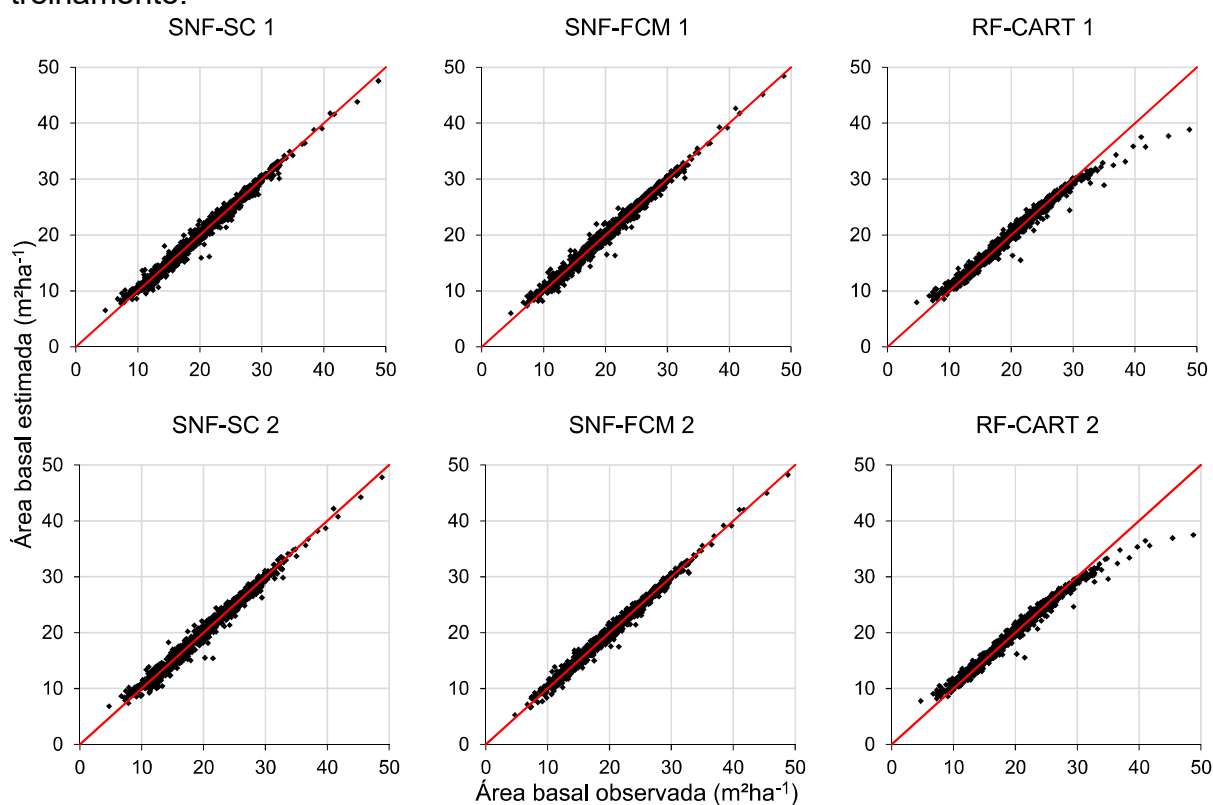
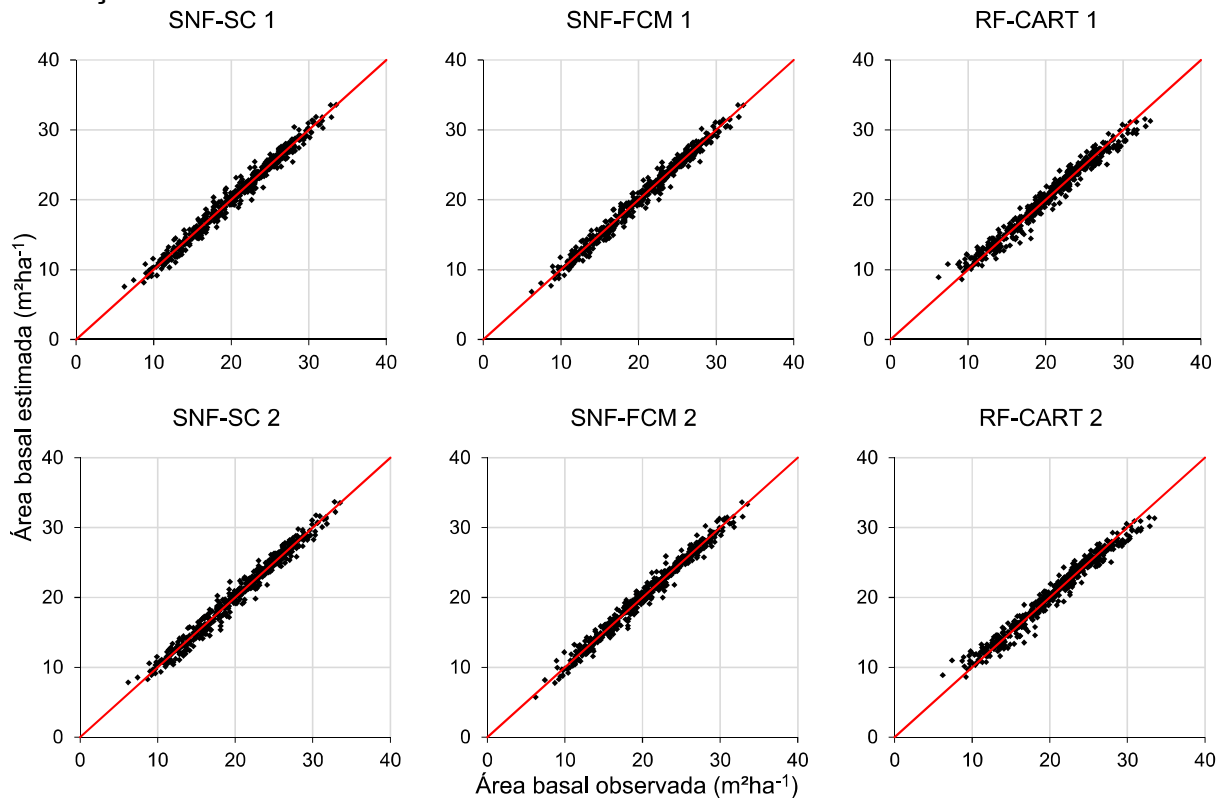




Figura 10 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) *versus* estimados na validação.



Os gráficos das Figuras 11 e 12, tanto no treino como na validação, em que foi empregado o SNF com os algoritmos SC e FCM (SNF-SC 1 e 2 e SNF-FCM 1 e 2), apresentaram melhor distribuição residual em torno da linha de erro zero quando comparada a RF (RF-CART 1 e 2). Em geral, observou-se que todos os gráficos do SNF e RF apresentaram distribuição dos erros com leve tendência a superestimação nos menores valores de área basal. Essa tendência também foi relatada por Castro et al. (2013) e Demolinari (2006). Segundo Vospernick et al. (2010), esse tipo de tendência é comum e ocorre em povoamentos com baixa densidade. Verificou-se ainda, com base na Figura 12, que a configuração SNF-FCM 1 apresentou a menor dispersão dos erros na validação, com maior concentração em  $\pm 10\%$ .

Figura 11 – Dispersão dos erros em porcentagem em relação aos valores estimados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) no treinamento.

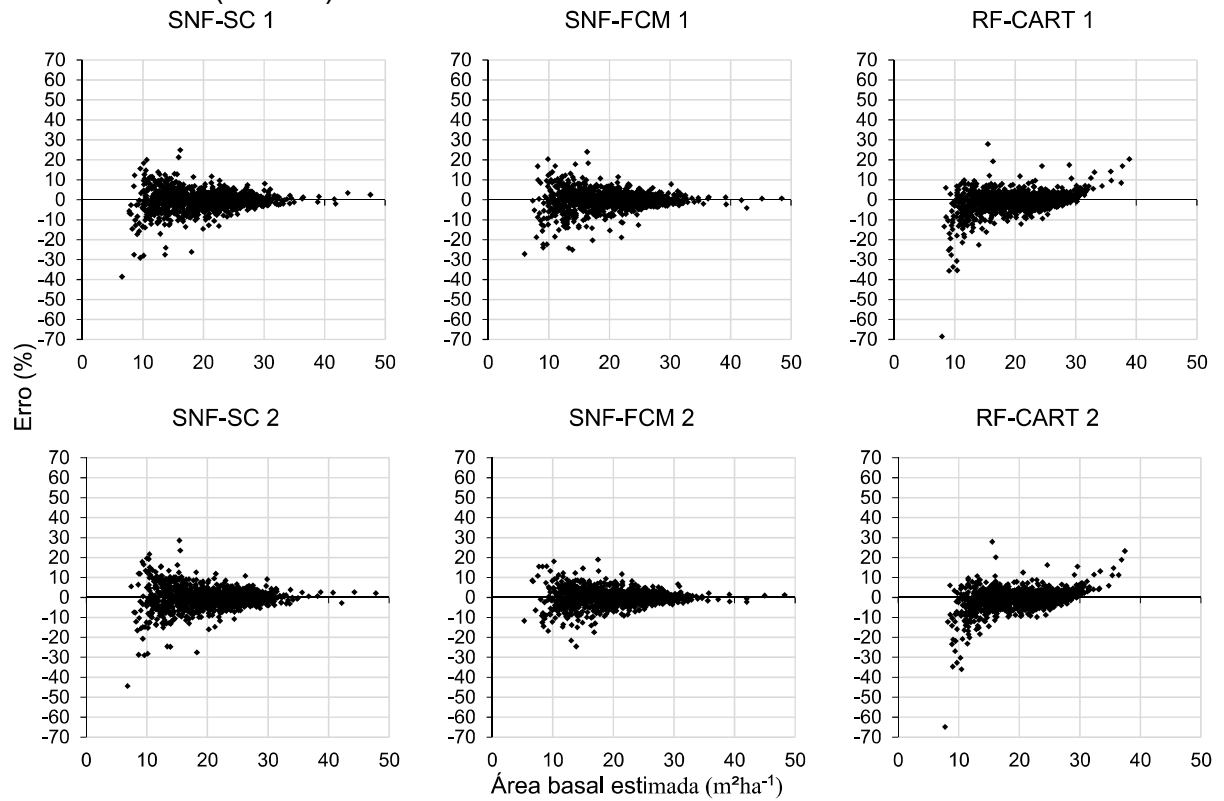
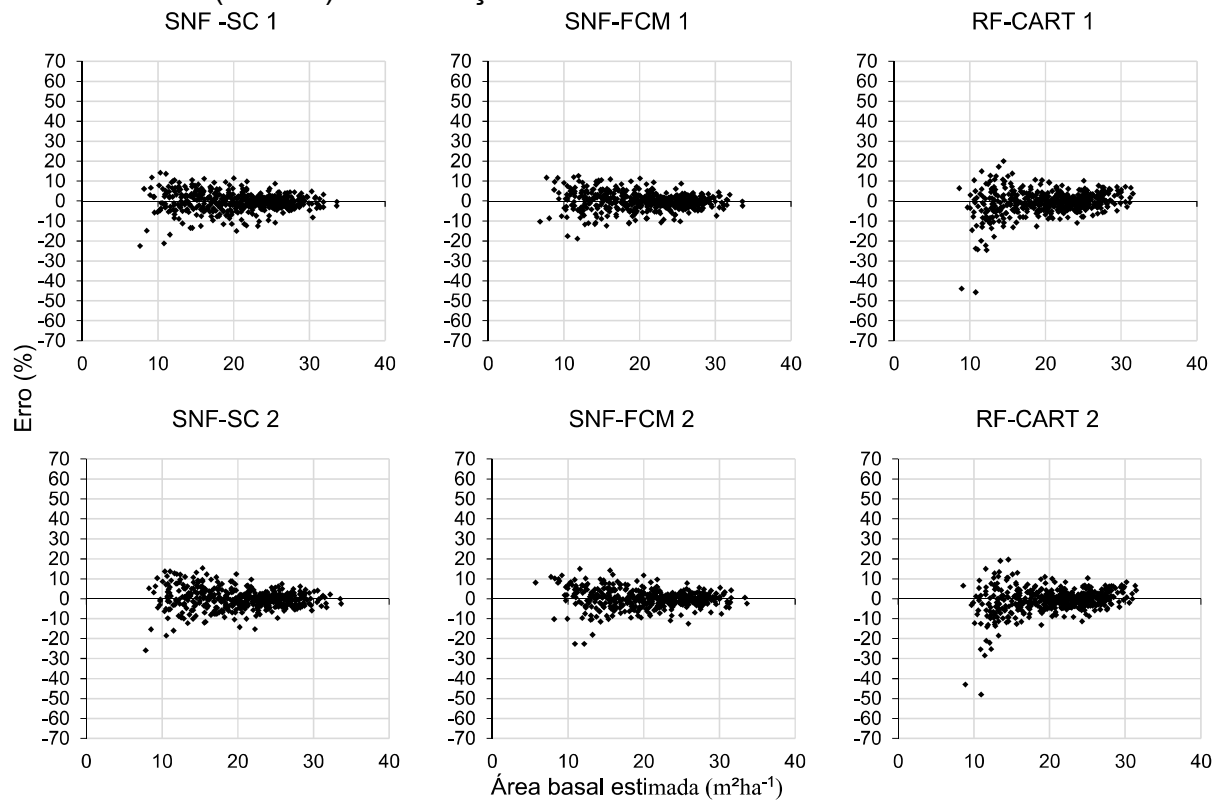


Figura 12 – Dispersão dos erros em porcentagem em relação aos valores estimados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) na validação.



Os gráficos das Figuras 9, 10, 11 e 12 da RF (RF-CART 1 e 2) demonstraram tendenciosidade mais acentuadas do que os do SNF, principalmente em superestimar as áreas basais menores que  $15 \text{ m}^2.\text{ha}^{-1}$  e subestimar as maiores que  $30 \text{ m}^2.\text{ha}^{-1}$ . Isto pode estar associado com mecanismos de obtenção das respostas numéricas no nó final da divisão da árvore, que é uma média dos valores observados de áreas basais, e esta é afetada pelos valores extremos. Segundo Zhang e Lu (2012), esse problema pode ser contornado ao calcular os resíduos e adicioná-los aos valores previstos. Basicamente, isso se faz treinando a RF para encontrar os valores previstos e calcular o erro de cada observação. Posteriormente, é treinada outra RF tendo o erro como variável resposta. O resultado final é soma dos valores previstos com os desvios estimados. Contudo, esse não foi o objetivo do presente trabalho.

Notou-se em todos os histogramas das Figuras 13 e 14, tanto no treinamento quanto na validação, que os erros em mais de 95% dos casos ficaram agrupados na amplitude das classes de  $\pm 10\%$ . Segundo Campos e Leite (2017), quanto mais fechada em torno de zero estiver a distribuição dos erros no histograma, melhor é o treinamento.

Figura 13 – Frequência relativa dos erros em porcentagem para estimativas da área basal ( $\text{m}^2.\text{ha}^{-1}$ ) no treinamento.

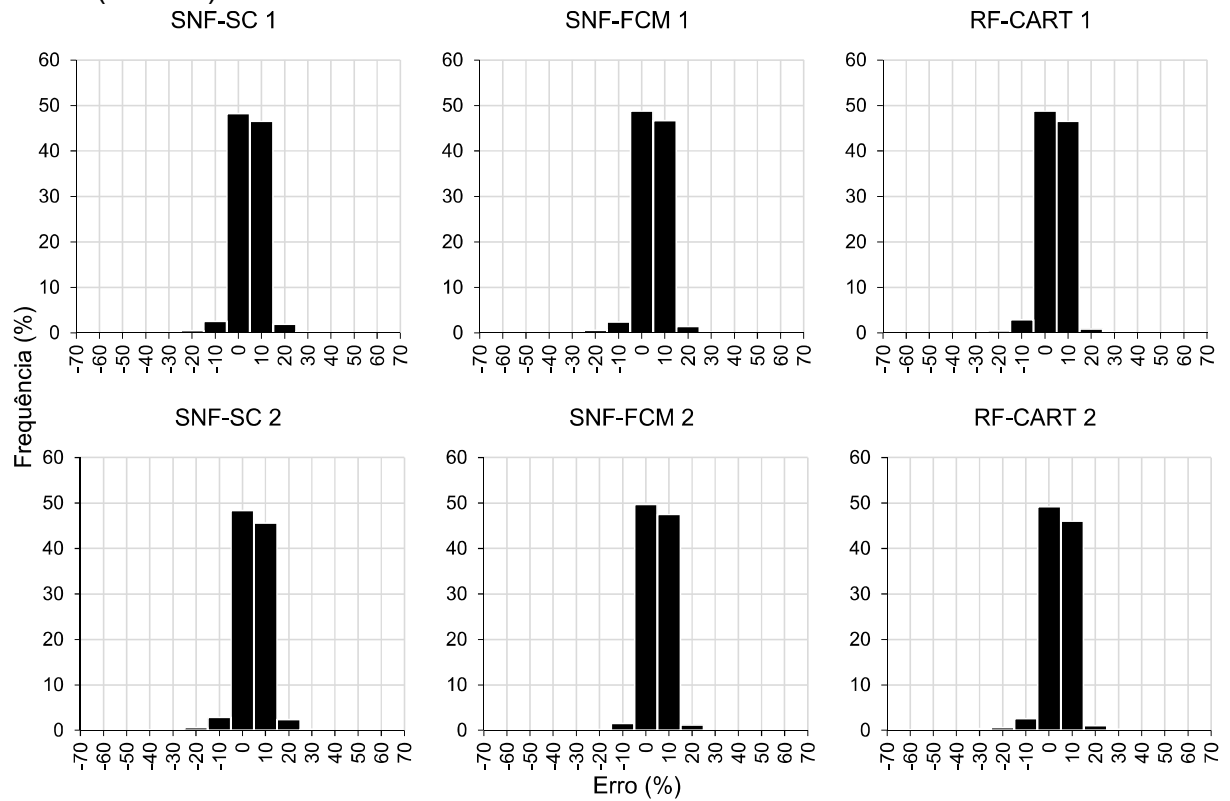
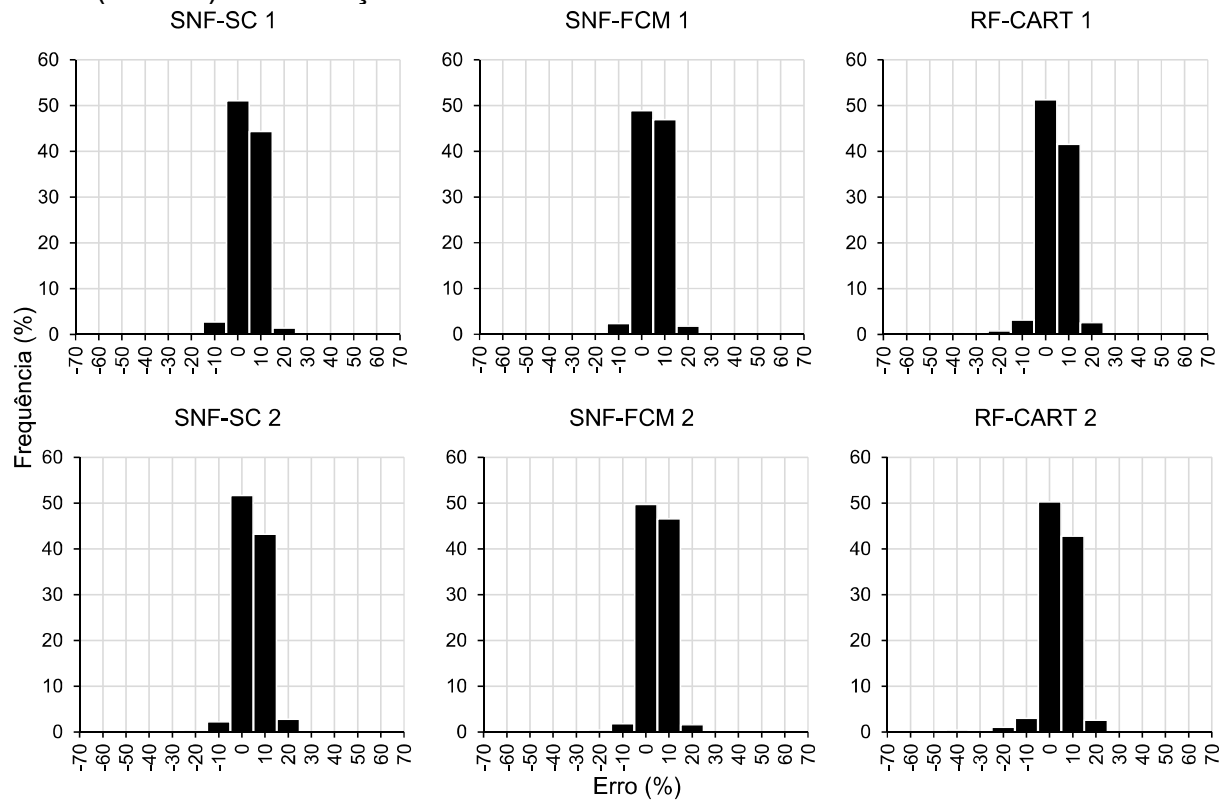


Figura 14 – Frequência relativa dos erros em porcentagem para estimativas da área basal ( $\text{m}^2.\text{ha}^{-1}$ ) na validação.



## 5. 2 Volume

### 5.2.1 Seleção das variáveis de entrada para prognose da produção volumétrica

De forma geral, todas as variáveis contribuem para estimação do volume futuro (Figura 15). Porém, pode ser notado na Figura 15B, que os valores de RMSE% para a validação aumentam e os de treinamento diminuem, conforme aumenta o número de entradas. Os resultados obtidos, indicam que o volume atual é a variável mais significativa para projeção do volume futuro em ambas as técnicas. Assim como aconteceu com a área basal futura, que o relevo e a textura são as variáveis com menos influência nas previsões (Figura 15A). Diferente do que ocorreu com a projeção de área basal futura, o clone foi mais importante do que as variáveis numéricas, idade atual e futura e área basal futura. Verifica-se ainda, que a combinação volume atual, área basal futura, idade futura e atual é a mais influente para prever o volume futuro. Estas variáveis mencionadas são as que apresentaram os maiores valores de índice de importância no RF. Os valores de RMSE de treinamento e validação foram de 5,15% e de 5,43%, respectivamente, para esta combinação.



### 5.2.2 Configurações das técnicas selecionadas para projetar o volume

As combinações das configurações geraram 833 SNF treinados com o algoritmo SC, 833 com o algoritmo FCM e 833 com a RF. Destas, foram pré-selecionadas as duas que obtiveram o menor erro de validação (Tabela 5).

Tabela 5 – Configurações pré-selecionadas do Sistema *Neuro-Fuzzy* e *Random Forest*

Técnica	Algoritmo	N°	Parâmetros	
			Fator squash	Raio
Sistema <i>Neuro-fuzzy</i> (SNF)	SC	3	1,50	0,30
	SC	4	1,25	0,40
			N° de cluster	Expoente
	FCM	3	16	1,215
	FCM	4	13	1,315
<i>Random Forest</i> (RF)			N° de árvores	N° de obs. por folha
	CART	3	400	10
	CART	4	350	15

em que: SC = *Subtractive Cluster*; FCM = *Fuzzy-C-Means*; CART = *Classification and regression Tree*.

As configurações selecionadas para projetar o volume foram denominadas de SNF-SC 3 (Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 3); SNF-SC 4 (Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 4); SNF-FCM 3 (Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 3); SNF-FCM 4 (Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 4); RF-CART 3 (*Random Forest* com algoritmo *Classification and regression Tree* para configuração 3) e RF-CART 4 (*Random Forest* com algoritmo *Classification and regression Tree* para configuração 4).

### 5.2.3 Desempenhos das configurações das técnicas de inteligência artificiais selecionadas para prognose do volume

Os algoritmos SC e FCM não demonstraram diferenças nas estatísticas para prognose do volume (Tabela 6), assim como ocorreu com a área basal (Tabela 4). A configuração SNF-FCM 3 teve as melhores estatísticas com relação as demais, apresentando o menor RMSE(%) e V(%) e o maior  $r_{\hat{Y}Y}$  nos conjuntos de treino e validação. Binoti et al. (2015), em seu trabalho com redes neurais para prognose da produção florestal, encontraram RMSE de 5,11% e 5,14% e coeficiente de correlação de 0,9932 e 0,9923 para o conjunto de treinamento e generalização, respectivamente. Martins et al. (2015), encontraram RMSE variando de 6,47% a 3,58% para o conjunto de treinamento trabalhando com diferentes configurações de redes neurais artificiais para prognosticar a produção de povoamentos clonais de eucalipto.

Tabela 6 – Estatísticas do treinamento e validação do Sistema *Neuro-Fuzzy* e *Random Forest* para projetar o volume ( $m^3.ha^{-1}$ )

Técnica- configuração	Treino			Validação		
	$r_{\hat{Y}Y}$	V(%)	RMSE (%)	$r_{\hat{Y}Y}$	V(%)	RMSE(%)
SNF-SC 3	0,9936	0,00	4,97	0,9919	0,13	5,39
SNF-SC 4	0,9932	0,00	5,11	0,9917	0,13	5,43
SNF-FCM 3	0,9941	0,00	4,77	0,9924	-0,01	5,18
SNF-FCM 4	0,9941	0,00	4,77	0,9923	0,05	5,21
RF-CART 3	0,9974	0,05	3,20	0,9902	-0,39	5,32
RF-CART 4	0,9973	0,03	3,21	0,9902	-0,35	5,42

em que: RMSE(%) = raiz do erro médio quadrático em porcentagem; v(%) = viés em porcentagem;  $r_{\hat{Y}Y}$  = coeficiente de correlação; SNF-SC 3 = Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 3; SNF-SC 4 = Sistema *Neuro-Fuzzy* com algoritmo *Subtractive Cluster* para configuração 4; SNF-FCM 3 = Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 3; SNF-FCM 4 = Sistema *Neuro-Fuzzy* com algoritmo *Fuzzy-C-Means* para configuração 4; RF-CART 3 = *Random Forest* com algoritmo *Classification and regression Tree* para configuração 3; RF-CART 4 = *Random Forest* com algoritmo *Classification and regression Tree* para configuração 4.

Os modelos da RF (RF-CART 3 e 4) apresentaram valores de RMSE (%) para o treinamento menores comparado à validação (Tabela 6). Possivelmente, pode ter ocorrido *overfitting*, que ocorre quando o modelo ajusta demais aos dados de treinamento e não consegue reproduzir bons resultados na validação. Segundo Souza (2017), espera-se maior exatidão não só para os dados de treinamento, mas também para os dados de validação. Dessa forma, é possível inferir que a técnica possui boa



generalização. O ideal é que os erros de treinamento e validação sejam próximos, como no SNF.

Nas Figuras 16 e 17, ambas as técnicas, tanto no treino como na validação, apresentaram a maioria dos dados distribuídos em torno da linha de 45°, o que confirma que os dados observados de volume estão fortemente relacionados com os valores estimados, como demonstrado pelos coeficientes de correlação na Tabela 6. Notou-se também que as configurações da RF (RF-CART 3 e 4) erraram mais nas áreas de maior produção em volume no conjunto de treinamento comparada as configurações do SNF (SNF-SC 3 e 4 e SNF-FCM 3 e 4). Nesse caso, a RF subestimou todos os volumes acima de 900 m<sup>3</sup>.ha<sup>-1</sup>. Este fato não foi notado nos gráficos de validação (Figura 15). Isto pode ter sido porque as áreas de produção têm volume baixo de 900 m<sup>3</sup>.ha<sup>-1</sup>.

Figura 16 – Valores observados de volume (m<sup>3</sup>.ha<sup>-1</sup>) *versus* estimados no treinamento.

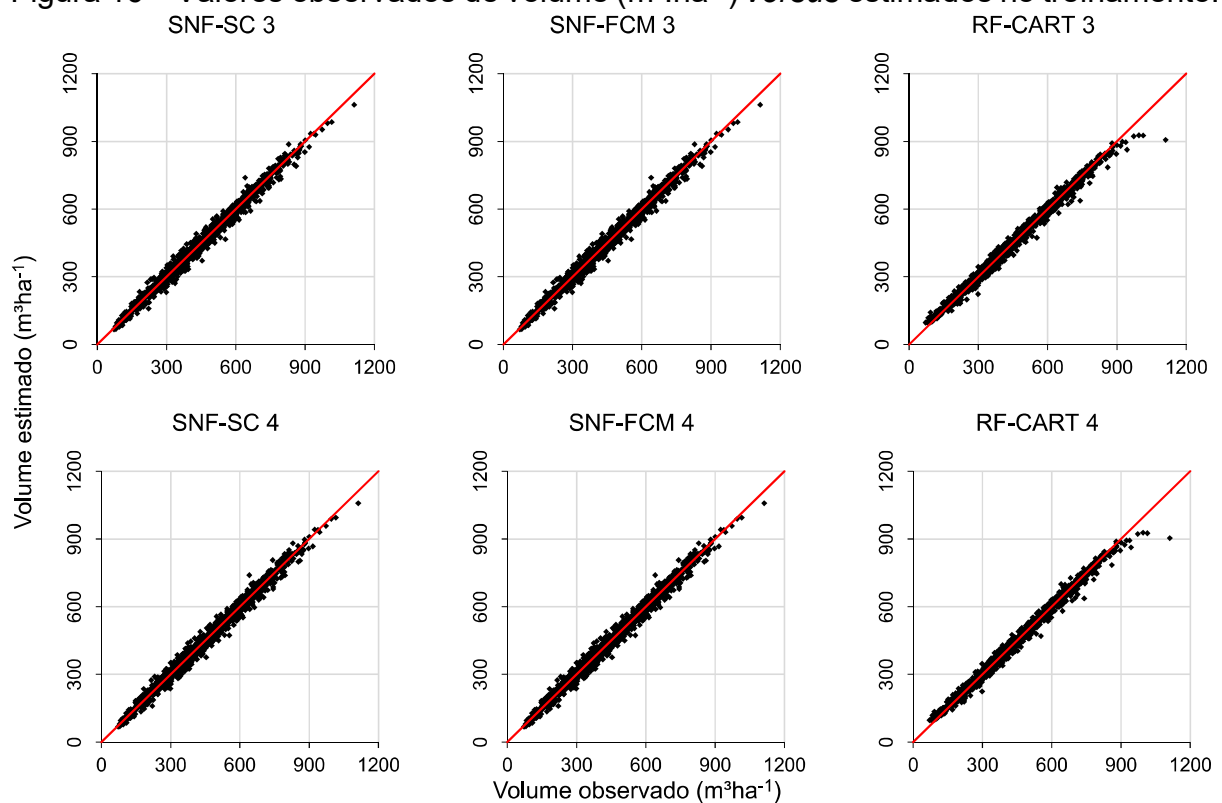
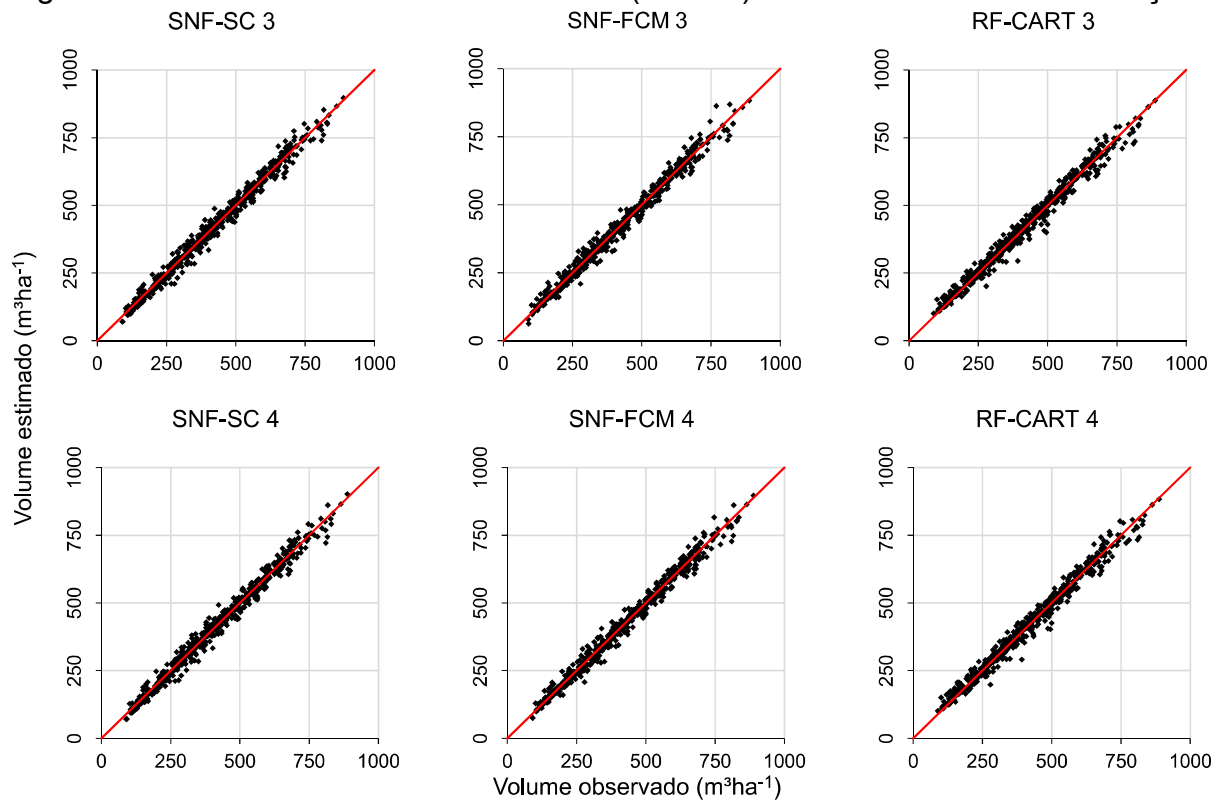


Figura 17 – Valores observados de volume ( $\text{m}^3 \cdot \text{ha}^{-1}$ ) *versus* estimados na validação.



A distribuição dos erros (Figuras 18 e 19), das configurações do SNF foram semelhantes, com erros variando de -30% a 30%, nos conjuntos de treino e validação, não evidenciando tendenciosidade. Em contraste, as configurações RF-CART 3 e 4, conforme apresentado nos gráficos de treinamento e validação (Figuras 18 e 19), apresentaram tendenciosidade em superestimar áreas com menores produções em volume e subestimar as áreas com os maiores volumes, como visto também na Figura 16.

Figura 18 – Dispersão dos erros em porcentagem em relação aos valores estimados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) no treinamento.

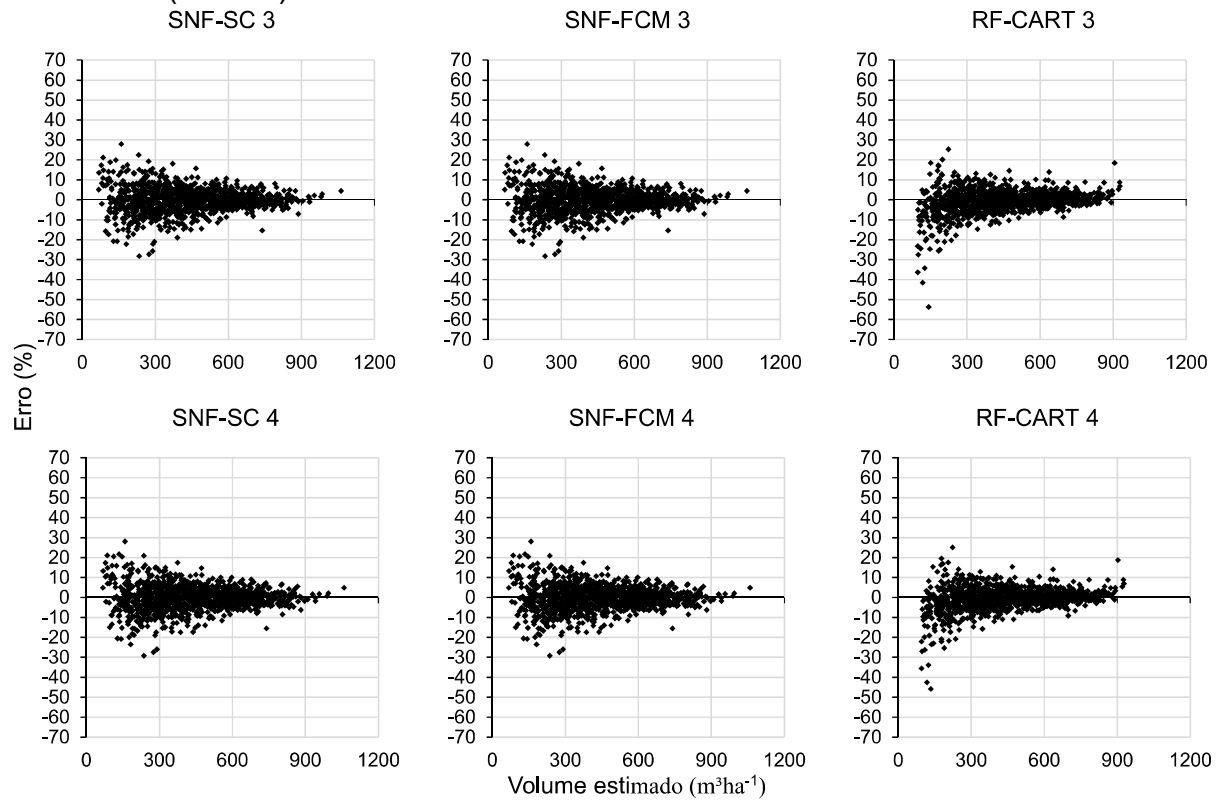
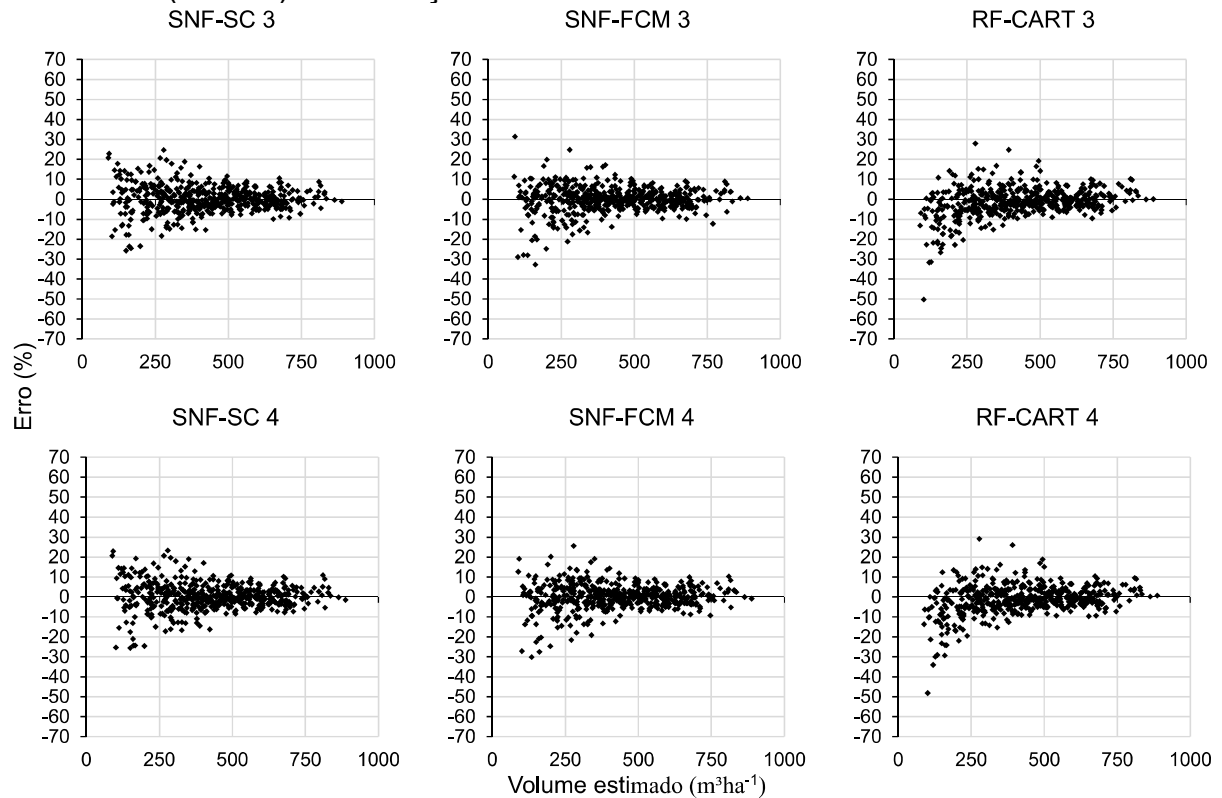


Figura 19 – Dispersão dos erros, em porcentagem, em relação aos valores estimados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) na validação.



A distribuição dos erros nos histogramas, para o conjunto de treinamento e validação, é próxima da distribuição normal. As duas técnicas tiveram cerca de 95% dos casos de erros nas classes de  $\pm 10\%$  (Figuras 20 e 21).

Figura 20 – Frequência relativa dos erros, em porcentagem, para estimativas de volume ( $\text{m}^3 \cdot \text{ha}^{-1}$ ) no treinamento.

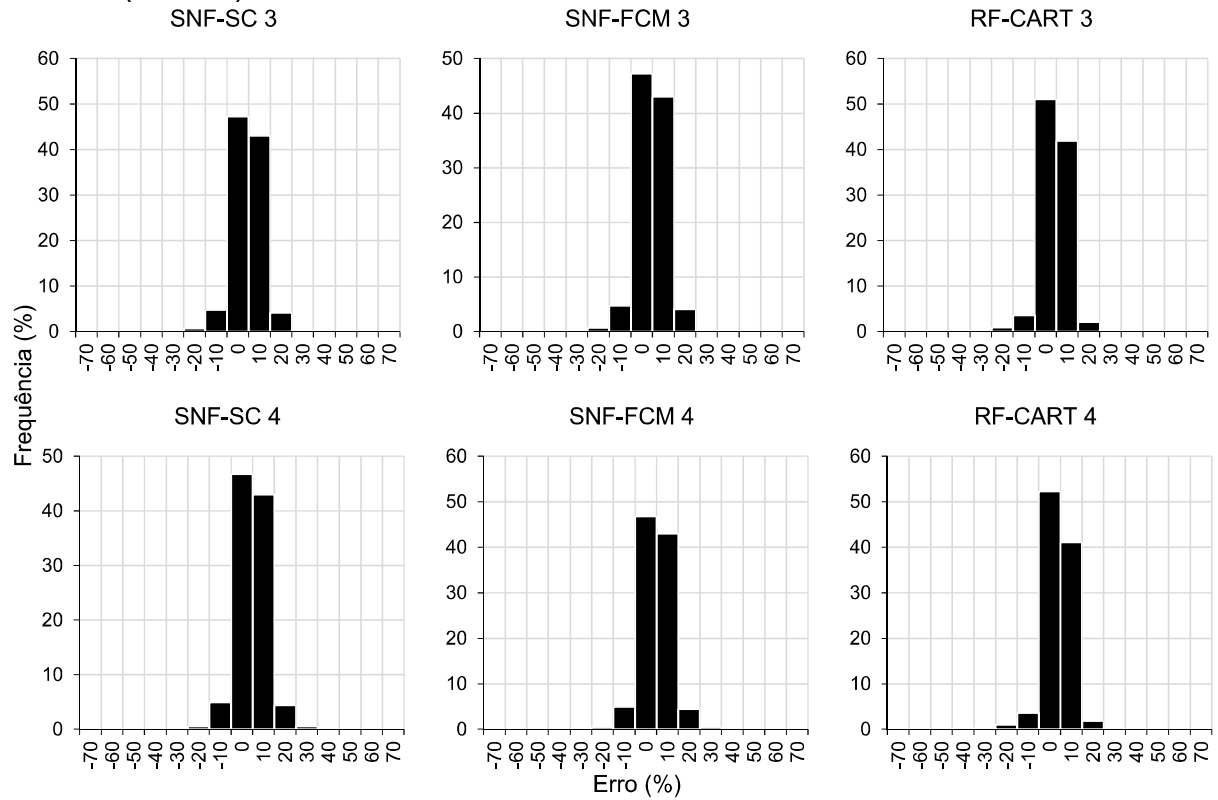
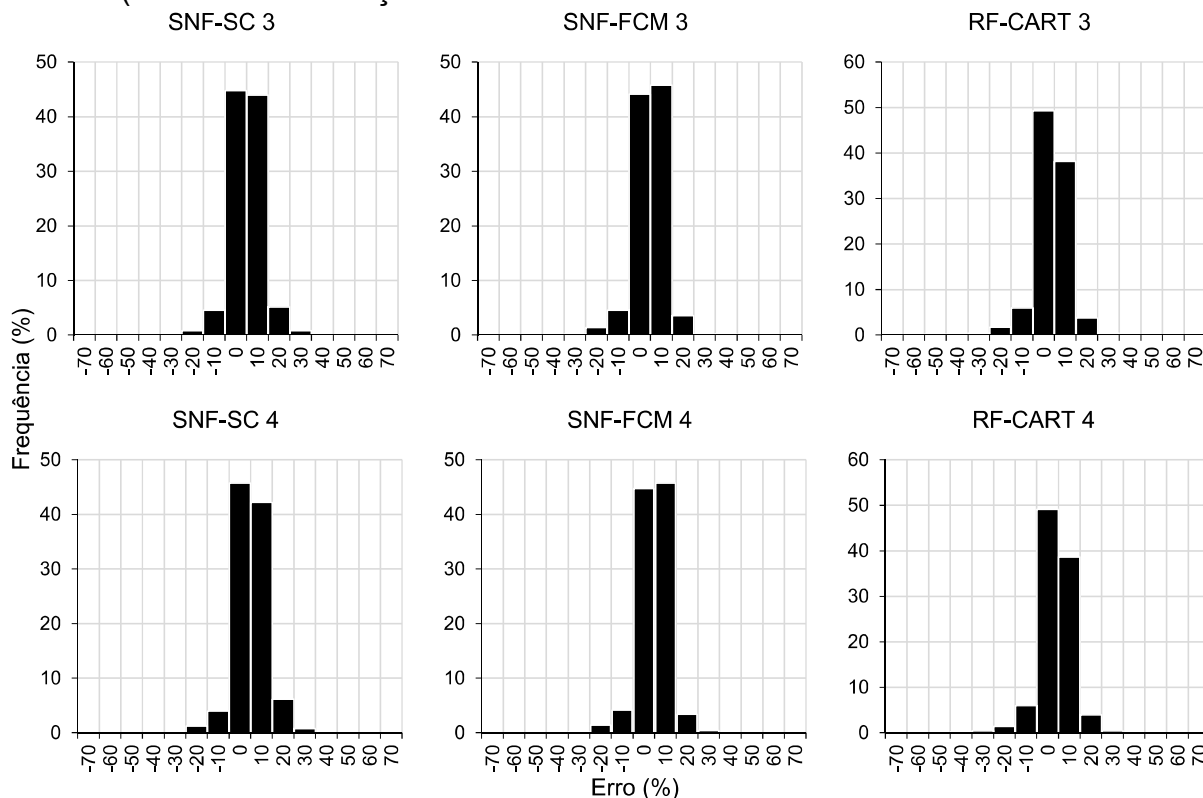


Figura 21 – Frequência relativa dos erros, em porcentagem, para estimativas de volume ( $\text{m}^3 \cdot \text{ha}^{-1}$ ) na validação.



### 5.3 Prognose da área basal

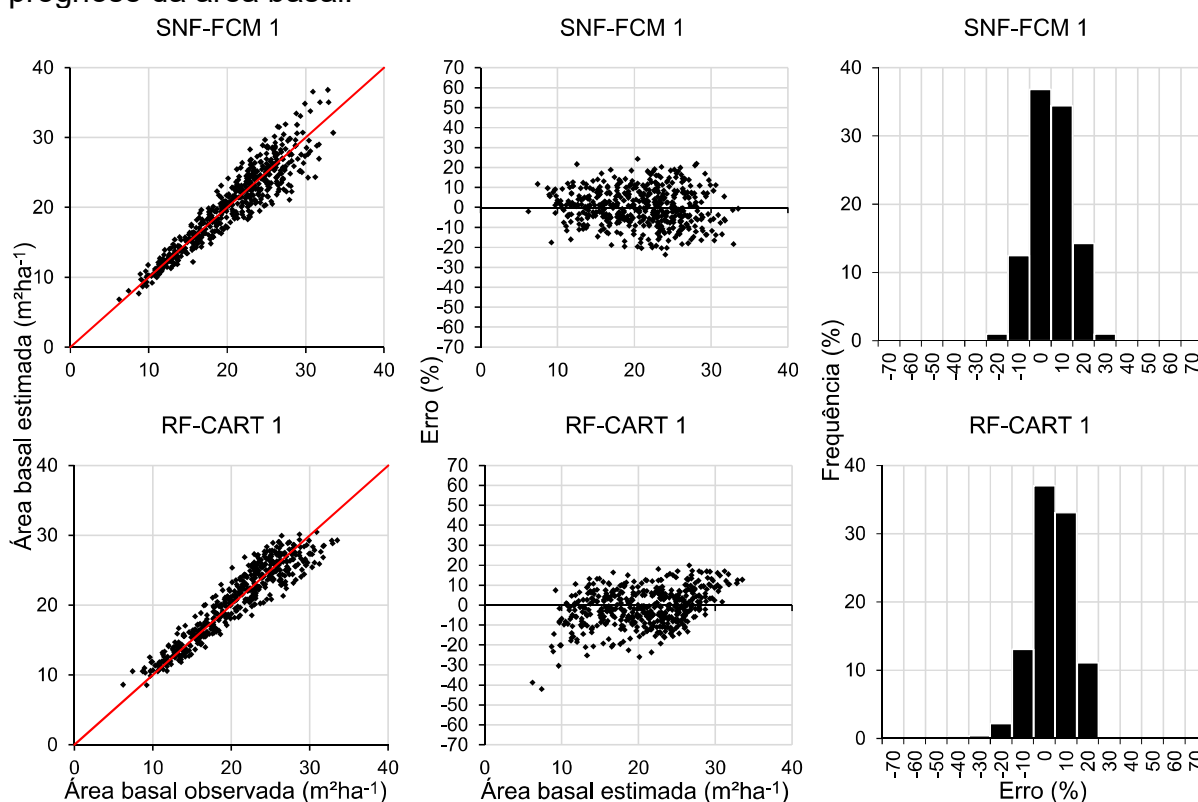
Ambas as técnicas forneceram bons resultados no treinamento e validação na prognose da área basal. As configurações do SNF tiveram estatísticas mais acuradas de RMSE, viés e coeficiente de correlação, bem como, apresentaram melhores distribuições residuais comparadas a RF. Dentre as configurações do SNF, não foi observado diferenças nas estatísticas (Tabela 4), sendo difícil a escolha da melhor configuração. Porém, a configuração SNF-FCM 1 apresentou pontos menos dispersos no gráfico de dispersão residual na validação comparada as demais, com valores de erro entre  $\pm 10\%$  (Figura 12), sendo esta a escolhida para projetar a área basal.

Considerando a RF e para fins comparativos com o SNF-FCM1, a configuração RF-CART 1 também foi escolhida para projetar a área basal. Cabe ressaltar que esta configuração apresentou menor RMSE e viés (Tabela 4) no conjunto de treinamento e validação com relação a RF-CART 2.

Para projetar a área basal, aplicou-se as configurações SNF-FCM1 e RF-CART 1 nos dados de validação do primeiro inventário florestal contínuo. As áreas basais futuras foram estimadas em projeções sucessivas, ou seja, ano a ano, usando como área basal atual o valor estimado do ano anterior, até a última medição de cada parcela.

Os valores de RMSE e  $r_{\hat{Y}\hat{Y}}$  na projeção da área basal para a configuração SNF-FCM 1 e RF-CART 1 foram de 9,2% e 9,17% e de 0,9416 e 0,9396, respectivamente. A perda de acurácia em comparação com as estatísticas de treinamento já era esperada, visto que, o erro da primeira projeção acumula com a segunda e assim sucessivamente até a última medição. Porém, não afetou de forma drástica as estimativas, já que as técnicas conseguiram modelar a tendência de crescimento em área basal, como mostrado na Figura 22.

Figura 22 – Valores observados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) *versus* estimados, dispersão dos erros em porcentagem em relação aos valores estimados de área basal ( $\text{m}^2.\text{ha}^{-1}$ ) e histograma da frequência relativa dos erros percentuais para os dados utilizados na prognose da área basal.



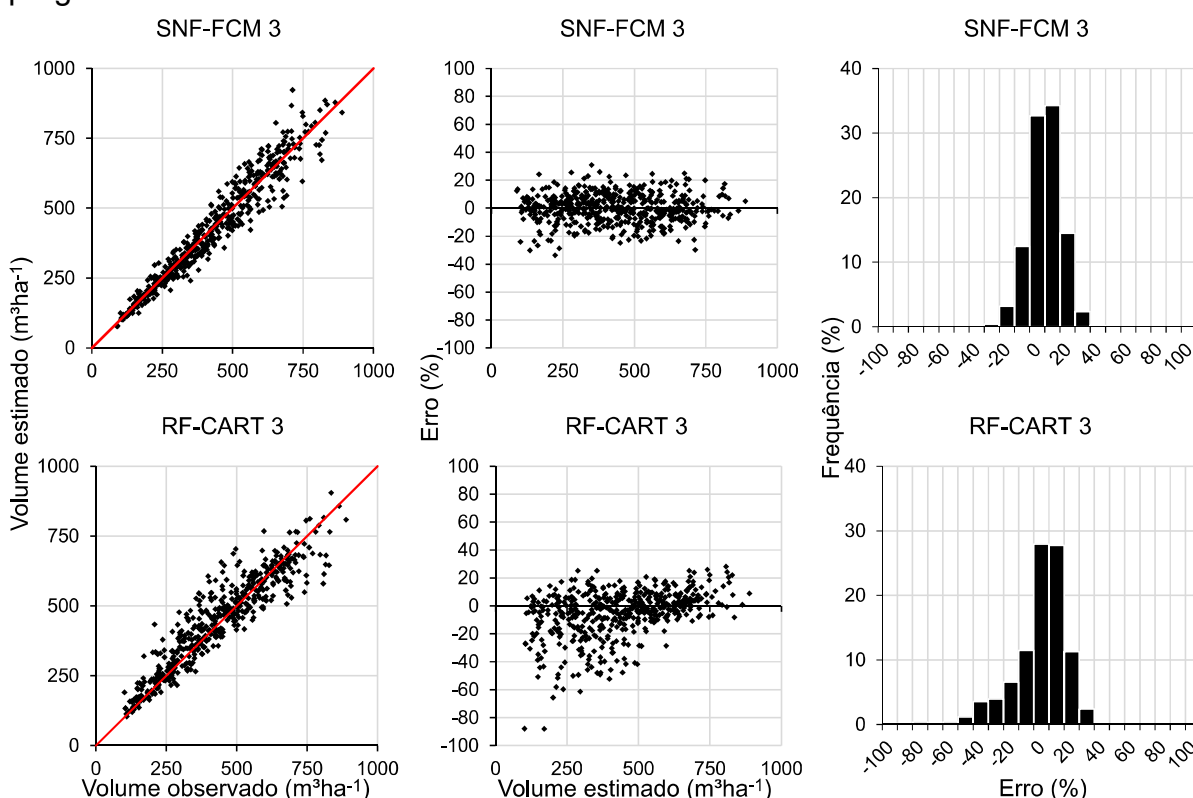
## 5.4 Prognose do volume

As configurações do SNF com os algoritmos SC e FCM foram semelhantes nas estatísticas e análise gráfica. Contudo, a configuração SNF-FCM 3 apresentou estatísticas mais exatas no treinamento e na validação (Tabela 6). Portanto, foi considerada a mais indicada para projeção do volume.

Assim como no SNF, as configurações da RF foram similares nas estatísticas e nas análises gráficas. Para a projeção do volume, foi escolhida a configuração RF-CART 3 por apresentar o menor RMSE (%) na validação comparada a RF-CART4 (Tabela 6).

Para projetar o volume, aplicou-se as configurações SNF-FCM 3 e RF-CART 4 nos dados de validação do primeiro inventário florestal contínuo. Cabe ressaltar que a área basal projetada foi utilizada como entrada nessa etapa. Os resultados estão apresentados na Figura 23.

Figura 23 – Valores observados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) *versus* estimados, dispersão dos erros em porcentagem em relação aos valores estimados de volume ( $\text{m}^3.\text{ha}^{-1}$ ) e histograma da frequência relativa dos erros percentuais para os dados utilizados na prognose do volume.



Percebeu-se que na Figura 23, que o gráfico do volume estimado *versus* observado, apresentou pontos mais dispersos em comparação aos gráficos de treinamento (Figura 16). Porém, observou-se no gráfico de dispersão dos erros percentuais que não houve tendência nas estimativas (Figura 23). Os erros percentuais ficaram em mais de 90% dos casos nas classes de -12% a +15%. Estes resultados são satisfatórios e comprovam o potencial do SNF na prognose da produção florestal.

Com relação a RF, a configuração RF-CART 3 apresentou pontos mais dispersos na linha de 45° no gráfico de volume estimado *versus* o observado comparada a configuração SNF-FCM 3 (Figura 23). Verificou-se ainda que houve tendência visível em superestimar as menores áreas de produção em volume e subestimar as de maiores. Pode ser notado também que o histograma apresentou uma leve assimetria a esquerda, com cerca de 70% dos casos de erros concentrados nas classes de -20% a +10%.

Os valores de RMSE e  $r_{\hat{Y}\hat{Y}}$  para projeção do volume utilizando as configurações SNF-FCM 3 e RF-CART 3 foram de 10,80% e 15,30% e de 0,9577 e 92,95, respectivamente. Notou-se perda de acurácia em comparação as estatísticas de treinamento, assim como ocorreu com a área basal. Binoti et al (2015), relataram este mesmo problema, utilizando diferentes tipos RNA para projetar o volume de povoamentos de eucalipto. Estes autores encontraram valores de RMSE e  $r_{\hat{Y}\hat{Y}}$  variando de 10,89% a 15,49% e de 0,9305 a 0,9652, respectivamente.

O SNF e RF conseguiram modelar a tendência de crescimento do povoamento clonal de eucalipto. Isto pode ser atribuído a dois fatores: o primeiro deles está relacionado a capacidade das técnicas em captar a não linearidade dos dados, o que faz delas, segundo Haykin (2001), aproximadores universais. Para Vieira et al. (2018), esta característica é importante na modelagem do crescimento e produção de povoamentos florestais.

O segundo fator pode estar associado ao emprego do material genético como variável de entrada. No trabalho de Binoti et al. (2015), o clone foi a variável mais importante dentre as categóricas para projetar o volume. Os autores, sugerem que isto é devido a especificidade do material genético às características ambientais. Vieira et al. (2018), afirmam que o clone sempre deve ser avaliado, já que o crescimento é diferente entre os mesmos. Segundo Husch (2002), o crescimento é



influenciado pela capacidade genética da espécie interagida com meio ambiente, como clima, solo, topografia e competição.

A modelagem do crescimento e produção tem sido amplamente estudada no setor florestal, com objetivo de fornecer informações nas tomadas de decisão e suporte para o gerenciamento das florestas. Este trabalho propôs duas abordagens novas para o cálculo de rendimentos futuros de área basal e volume, em nível de povoamento total.

Como na modelagem dos recursos florestais muitas vezes apresentam variáveis qualitativas e com relações complexas, o SNF e RF, com os resultados apresentados neste trabalho, apresentam-se como alternativas de modelagem. As técnicas permitem boa aproximação de sistemas não-lineares. Em termos de desempenho e acurácia, tanto o algoritmo SC como o FCM fornecem bons resultados. O benefício de utilizar esses algoritmos para gerar o sistema de inferência *fuzzy* é a redução no número de regras, principalmente se o número de variáveis de entrada for muito elevado, visto que, o valor aumenta exponencialmente com o número de funções de associação (WEI et al., 2007).

Com relação a *Random Forest*, o seu uso deve ser avaliado com cuidado, principalmente quando existirem valores da variável dependente com alta amplitude. Isso porque, no seu mecanismo de funcionamento, a obtenção da resposta se dá pela média dos valores observados. Nesse caso, poderá acarretar superestimativas nas menores classes e subestimativas nas maiores. Entretanto, a técnica demonstrou bom desempenho para prognose da área basal e do volume, além de fornecer as variáveis que melhor contribuem para a previsão.

Para construir um SNF para prognose da produção florestal, é necessária identificação de variáveis mais relevantes na projeção da área basal e volume. Com os resultados apresentados nesse trabalho, pode-se observar que a inclusão de variáveis irrelevantes causa perda de generalização, além de aumentar a complexidade do SNF, bem como, o tempo de processamento. O método de seleção de variáveis do SNF com algoritmo de mínimos quadrados é viável, visto que, reduziu o número de variáveis e melhorou os resultados de generalização. Percebeu-se ainda, que as variáveis selecionadas foram as mesmas entre as duas técnicas avaliadas.

A exatidão das estimativas na produção florestal proporcionadas pelas técnicas SNF e RF depende da calibração de suas configurações. Deve-se dar preferência a

configurações mais simples (BRAGA et al. 2007), desde que estas consiga captar as relações entre os dados de entrada e saída.

Tanto o SNF quanto a RF possuem grande potencial para serem utilizadas como ferramentas auxiliares no manejo de plantios florestais, possibilitando prever os fatores que contribuem para a produtividade como clone, espaçamento, tipo de solo, relevo e textura do solo.

## 6 CONCLUSÕES

Em geral, todas as variáveis numéricas foram influentes na projeção da área basal e volume. Dentre as variáveis categóricas, o clone foi a variável que mais influenciou. Já o relevo e a textura foram as menos influentes.

As configurações dos algoritmos *Fuzzy-C-Means* e *Subtractive Cluster* apresentaram resultados estatísticos semelhantes para prognose da área basal e do volume. O Sistema *Neuro-Fuzzy* apresentou maior exatidão nas estimativas geradas para as variáveis citadas acima quando comparado ao *Random Forest*.

O Sistema *Neuro-Fuzzy* e *Random Forest* apresentaram resultados satisfatórios para prognose da área basal e do volume, confirmando a aplicabilidade dessas técnicas na modelagem da produção florestal.

## REFERÊNCIAS

ABREU, E. C. R et al. Modelagem para prognose precoce do volume por classe diamétrica para *Eucalyptus grandis*. **Scientia Forestalis**, n.61, p.86-102, 2002.

AKBARZADEH, S. et al. Prediction of Conductivity by Adaptive Neuro-Fuzzy Model. **PLoS ONE**, v. 9, n. 3, p. e92241, 2014.

AKKOÇ, B; ARSLAN, A; KÖK, H. Automatic gender determination from 3D digital maxillary tooth plaster models based on the random forest algorithm and discrete cosine transform. **Computer Methods and Programs in Biomedicine**, v. 143, p. 59-65, 2017.

ARAÚJO, A. G. **Estimativa e classificação da compactação do solo pelo tráfego de máquinas agrícolas através da modelagem nebulosa**. 2004. 241 p. Tese (Doutorado em Engenharia Elétrica) - Escola Politécnica da Universidade de São Paulo. São Paulo, 2004.

ARCHER, K. J.; KIMES, R. V. Empirical characterization of random forest variable importance measures. **Computational Statistics & Data Analysis**, v. 52, n. 4, p. 2249-2260, 2008.

ASHRAF, M. I. et al. A novel modelling approach for predicting forest growth and yield under climate change. **PloS One**, v. 10, n. 7, p. e0132066, 2015.

BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms**. Springer Science & Business Media, 1981.

BILGEHAN, M. Comparison of ANFIS and NN models: with a study in critical buckling load estimation. **Applied Soft Computing**, v. 11, n. 4, p. 3779–3791, 2011.

BINOTI, M. L. M. da S. et al. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, v. 21, n. 1, p. 97-105, 2015.

BINOTI, M. L. M. da S. et al. Redes neurais artificiais para estimação do volume de árvores. **Revista Árvore**, v. 38, n. 2, p. 283-288, 2014.

BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639-645, 2013.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Editora LTC, 2007. 262p.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

CAMPOS, J. C. C; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 3. ed. Viçosa, MG: UFV, 2017. 548 p.

CANO, G. et al. Automatic selection of molecular descriptors using random forest: Application to drug discovery. **Expert Systems with Applications**, v. 72, p. 151-159, abr. 2017.

CARVALHO, S. P. C. **Uma nova metodologia de avaliação do crescimento e da produção de *Eucalyptus* sp. clonal para fins energéticos**. UFLA: Lavras, 2010. 103 p. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras, 2010.

CASTRO, R. V. O. et al. Modelagem do crescimento e produção para um povoamento de *Eucalyptus* utilizando dois métodos para quantificação do índice de local. **Scientia Forestalis**, v.43, n.105, p.83-90, 2015.

CASTRO, R. V. O. et al. Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos. **Pesquisa Agropecuária Brasileira**, v. 48, n. 3, p. 287-295, 2013.

CASTRO, R. V. O. et al. Individual growth model for eucalyptus stands in Brazil using artificial neural network. **ISRN Forestry**, v. 2013, 2013.

CHAVES, L. E. **Modelos computacionais fuzzy e neuro-fuzzy para avaliarem os efeitos da poluição do ar**. 2013. 113 p. Tese (Doutor em Engenharia Mecânica na área de Projetos) – Universidade Estadual Paulista, Guaratinguetá, 2013.

CHELGANI, S. C; MATIN, S. S.; MAKAREMI, S. Modeling of free swelling index based on variable importance measurements of parent coal properties by random forest method. **Measurement**, v. 94, p. 416-422, 2016.

CHERRY et al. Sequential Monte Carlo tracking of the marginal artery by multiple cue fusion and random forest regression. **Medical image analysis**, v. 19, n. 1, p. 164-175, 2015.

CHIU, S. Fuzzy Model Identification Based on Cluster Estimation. **Journal of Intelligent & Fuzzy Systems**, v. 2, n. 3, p. 267-278, 1994.

CLUTTER, J. L. et al. **Timber management**: a quantitative approach. New York: John Wiley and Sons, 1983. 333 p.

CRACKNELL, M. J.; READING, A. M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. **Computers & Geosciences**, v. 63, p. 22-33, 2014.

DAVIS, L. S.; JOHNSON, K. N. **Forest management**. 3.ed, New York: McGraw-Hill, 2000. 790p.

DEMOLINARI, R. de A. et al. Crescimento de plantios clonais de eucalipto não desbastados na região de Monte Dourado (PA). **Revista Árvore**, v.31 p.503-512, 2007.

DIAMANTOPOULOU, M. J. Predicting fir trees stem diameters using Artificial Neural Network models. **Southern African Forestry Journal**, v. 205, n. 1, p. 39-44, 2005.

FALLAHIAN S.; SEYEDPOOR, S. M. A two stage method for structural damage identification using an adaptive neuro-fuzzy inference system and particle swarm optimization. **Asian Journal of Civil Engineering**, v.11, n. 6, p. 795-808, 2010.

GOUDARZI, N. Free variable selection QSPR study to predict <sup>19</sup>F chemical shifts of some fluorinated organic compounds using Random Forest and RBF-PLS methods. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 158, p. 60-64, 2016.

GUIMARÃES, D. P. **Desenvolvimento de um modelo de distribuição diamétrica de passo invariante para prognose e projeção da estrutura de povoamento de eucalipto**. 1994. 178p. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa.

GRĄBCZEWSKI, K. **Meta-Learning in Decision Tree Induction**. In: Studies in Computational Intelligence. Cham: Springer International Publishing, 2014. 349p.

HASENAUER, H. Princípios para a modelagem de ecossistemas florestais. **Ciência e Ambiente**, Santa Maria, v. 20, n. 1, p. 53-69, 2000.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**, ed. 2. J. Edition. New York: Springer, 2009.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900p.

HUSCH, Bertram; BEERS, Thomas W.; KERSHAW JR, John A. **Forest mensuration**. John Wiley & Sons, 2002.

IBRAHIM, I. A.; KHATIB, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. **Energy Conversion and Management**, v. 138, p. 413–425, 2017.

INDÚSTRIA BRASILEIRA DE ÁRVORES (IBÁ). **Relatório IBÁ 2017**. Brasília, 2017. 80 p.

JAMES, G. et al. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. 426 p.

JANG, J.-S. R. Anfis: Adaptive-network-based fuzzy inference systems. **IEEE Transactions on Systems, Man, and Cybernetics**, [S.l.], v. 23, p. 714-723, May-June 1993.

JANG, J.-S. R. Input selection for ANFIS learning. In: **Fuzzy Systems, 1996. Proceedings of the Fifth IEEE International Conference on**. IEEE, 1996. p. 1493-1499.

JANITZA, S; TUTZ, G; BOULESTEIX, A-L. Random forest for ordinal responses: prediction and variable selection. **Computational Statistics & Data Analysis**, v. 96, p. 57-73, 2016.

PAL, S., MITRA, P. **Case Generation: A Rough-fuzzy Approach**, Workshop em Soft Computing in Case-Based Reasoning, International Conference on Case-Based Reasoning, ICCBR'01 Vancouver, British Columbia, Canada. 1992.

JEONG, J. H. et al. Random Forests for Global and Regional Crop Yield Predictions. **PLOS ONE**, v. 11, n. 6, p. e0156571, 2016.

MARTINS, F. B. **Modelagem de crescimento em nível de árvore individual para plantios comerciais de eucaliptos**. 2011. 159p. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa.

MASOUMI, F. et al. Integration of spectral, thermal, and textural features of ASTER data using Random Forests classification for lithological mapping. **Journal of African Earth Sciences**, v. 129, p. 445-457, 2017.

MASSELLI, Y. M. C. **Proposta de um modelo de redes neuro-fuzzy-polinomial otimizado por algoritmos de enxame aplicados à previsão**. 2009. 140 p. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Itajubá, Itajubá, 2009.

MATHUR, N.; GLESK, I.; BUIS, A. Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses. **Medical Engineering & Physics**, v. 38, n. 10, p. 1083-1089, 2016.

MATHWORKS, INC. **Fuzzy logic toolbox: for use with MATLAB®: user's guide**. MathWorks, 2017.

MIRANDA, D. L. C. de; JUNIOR, V. B.; GOUVEIA, D. M. Fator de forma e equações de volume para estimativa volumétrica de árvores em plantio de *Eucalyptus urograndis*. **Scientia Plena**, v. 11, n. 3, p. 1-8, 2015.

MOHAMMADI, Kasra et al. Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. **Renewable and Sustainable Energy Reviews**, v. 63, p. 423-434, 2016.

NUNES, M. H.; GÖRGENS, E. B. Artificial Intelligence Procedures for Tree Taper Estimation within a Complex Vegetation Mosaic in Brazil. **PloS one**, v. 11, n. 5, p. e0154738, 2016.

REIS, L. P. **Modelagem de floresta inequidiana: redes neurais artificiais aplicadas em uma floresta manejada no leste da Amazônia**. 2017. 91p. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa.

REIS, L. P. et al. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. **Forest Ecology and Management**, v. 382, p. 161-167, 2016.

SCOLFORO, J. R. S. **Biometria florestal**: Crescimento e produção florestal. Lavras: UFLA/FAEPE, 2006. 352 p.

SHAIKHINA, T. et al. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. **Biomedical Signal Processing and Control**, p. 1-14, 2017.

SILVA et al. Fuzzy and Neuro-Fuzzy Modeling for Total Volume study of Eucalyptus sp. In: **Hybrid Intelligent Systems, 2007. HIS 2007. 7th International Conference on**. IEEE, 2007. p. 358-361.

SILVA, R. M. et al. Fuzzy and neuro-fuzzy estimates of the total height of eucalyptus trees. In: **Proceedings of the 2008 ACM symposium on Applied computing**. ACM, 2008. p. 1772-1776.

SIMINSKI, K. Interval type-2 neuro-fuzzy system with implication-based inference mechanism. **Expert Systems with Applications**, v. 79, p. 140-152, 2017.

SIMÕES, M. G.; SHAW, I. S. **Controle e Modelagem Fuzzy**. 2. ed. São Paulo: Blucher. FAPESP, 2007. 186 p.

SOUZA, L. A. **Estimação de altura, volume e afilamento de arvores de eucalipto utilizando máquina de vetor de suporte**. 2017. 89p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES.

SU, S; ZHAO, S. An optimal clustering mechanism based on Fuzzy-C means for wireless sensor networks. **Sustainable Computing: Informatics and Systems**, 2017.

TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its application to modeling and control. **IEEE Trans. On Systems, Man & Cybernetics**, n. 15, p. 116-132, 1985.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining**: mineração de dados. Rio de Janeiro: Editora ciências moderna Ltda, 2009. 900 p.

TOOKE, T. R.; COOPS, N. C.; WEBSTER, J. Predicting building ages from LiDAR data with random forests for building energy modeling. **Energy and Buildings**, v. 68, n. PARTA, p. 603-610, 2014.

VANCLAY, J. K. **Modelling forest growth and yield**: applications to Mixed Tropical Forest. Wallingford: CAB INTERNATIONAL, 1994. 312 p.

VIEIRA, G. C. et al. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science of the Total Environment**, v. 619, p. 1473-1481, 2018.



VOSPERNIK, S.; MONSERUD, R. A.; STERBA, H. Do individual-tree growth models correctly represent height: diameter ratios of Norway spruce and Scots pine?. **Forest ecology and management**, v. 260, n. 10, p. 1735-1753, 2010.

WANG, L. et al. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. **The Crop Journal**, v. 4, n. 3, p. 212–219, 2016.

WEI, M. et al. Predicting injection profiles using ANFIS. **Information Sciences**, v. 177, n. 20, p. 4445-4461, 2007.

YU, X. et al. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 1, p. 28-37, 2011.

ZAMO, M. et al. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. **Solar Energy**, v. 105, p. 792-803, 2014.

ZADEH, L. A. Fuzzy Sets. **Information and Control**, v. 8; n. 3, p. 338-353, June 1965.

ZHANG, G; LU, Y. Bias-corrected random forests in regression. **Journal of Applied Statistics**, v. 39, n. 1, p. 151-160, 2012.

ZHANG, S.; AMATEIS, R. L.; BURKHART, H. E. Constraining individual tree diameter increment and survival models for loblolly pine plantations. **Forest Science**, Bethesda, v. 43, n. 3, p. 414-423, 1997.